Developing tourism demand forecasting models using machine learning techniques with trend, seasonal, and cyclic components

S. Cankurt and A. Subasi

Abstract—**This paper proposes the deterministic generation of auxiliary variables, which outline the seasonal, cyclic and trend components of the time series associated with tourism demand for the machine learning models. To test the contribution of the deterministically generated auxiliary variables, we have employed multilayer perceptron (MLP) regression, and support vector regression (SVR) models, which are the well-known state-of- art machine learning models. These models are used to make multivariate tourism forecasting for Turkey respected to two data sets: raw data set and data set with deterministically generated auxiliary variables. The forecasting performances are compared regards to these two data sets. In terms of relative absolute error (RAE) and root relative squared error (RRSE) measurements, the proposed machine learning models have achieved significantly better forecasting accuracy when the auxiliary variables have been employed.**

Index Terms—**Tourism Demand Forecasting, Multivariate Forecasting, Seasonal Time Series, Auxiliary Variable, Artificial Neural Network, Multilayer Perceptron, Support Vector Regression.**

I. INTRODUCTION

IN tourism demand forecasting, the quantitative forecasting models can be classified into three broad categories: time-series, econometric techniques, and artificial intelligence (AI) methods and its sub or interrelated disciplines like soft computing, machine learning, and data mining. Intelligence techniques which are predominantly derived from artificial neural network (ANN), support vector machine, fuzzy logic, genetic algorithms, swarm intelligence, have emerged in the tourism forecasting literature [1].

Forecasting situations vary widely in their time horizons involved, on how well we selected the factors that contribute the relationship in the model, types of data patterns, and many other aspects [2].

Most of the tourism demand time series very often exhibit patterns in terms of the seasonal, cyclic and trend components. Seasonality is one of the most important patterns, which is mostly observed in tourism demand time series. Therefore, it is one of the considerations to develop time series models. Some of these time series models are autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA), Holt-Winters models. These traditional statistical models require eliminating the effect of seasonality from a time series before making forecasting [3]. For example, Holt's model is developed for the time series in which there is neither trend nor seasonality, Brown's exponential smoothing model is suitable for time series in which there is a linear trend but no seasonality and Winters' additive and multiplicative models work well for time series both with a linear trend and a seasonal effect.

In tourism demand analysis, dealing with the seasonal fluctuations of tourism data has always been an important and complex issue. In the tourism demand forecasting taxonomy, seasonality can be faced either as a deterministic component or a stochastic component in the time series. If seasonality is considered as stochastic, we may handle seasonality by using deseasonalization or differencing approaches to make the seasonal adjustment and to remove the seasonal effect from the data. Then, we are able to apply these traditional statistical models to time series with seasonal component. If the seasonality is regarded as deterministic, introducing seasonal dummy variables into the time-series models would be sufficient to deal with the seasonal variations [1].

On the other hand, machine learning, which is an important area of artificial intelligence, has been successfully applied to many forecasting applications including the tourism demand forecasting. Artificial neural network (ANN) and support vector machine (SVM) are the most widely used state-of-art machine learning methods for forecasting. These models have been immerged in the field of the tourism demand forecasting. In different studies [3], ability of machine learning models in recognizing and learning the seasonal patterns without

S. CANKURT is with the Department of Information Technologies, International Burch University, Sarajevo, Bosnia and Herzegovina (e-mail: selcuk.cankurt@ibu.edu.ba).

A. SUBASI is with the Department of Information Technologies, International Burch University, Sarajevo, Bosnia and Herzegovina (e-mail: asubasi@ibu.edu.ba).

removing them from the raw data is shown. However, in the literature, effects of the seasonal, trend and cyclic patterns are not discussed for the machine learning models as much investigated as for the time series models.

Nevertheless, there are some attempts to develop techniques to deal with the patterns in the use of machine learning models in order to improve the accuracy of models. For example, in the study held by Hill et al. [4], a concept of deseasonalizing seasonal time series is used. They have used the deseasonalized data generated by the moving-to-ratio-average method to input a neural network model. Wang and Leu [5] proposed a hybrid model, which is a recurrent neural network trained by features extracted from ARIMA analyses, to forecast the mid-term price trend of the Taiwan stock exchange. That empirical study showed that the neural network trained by differenced data achieved better predictions than one trained by raw data [6].

Tseng et al. proposed a hybrid forecasting model, which combines the seasonal time series ARIMA (SARIMA) and the neural network back propagation (BP) models to forecast time series with seasonality [6]. They have used Hill et al. [4]'s concept of deseasonalizing a seasonal time series to formulate a neural network model and Wang and Leu [5]'s approach of taking the differences of a seasonal time series data to build another neural network model. In their approaches, they have trained a neural network with deseasonalized data (input the deseasonalized data generated by the moving-to-ratio-average method to the input layer) and differenced data (input the differenced data generated by the SARIMA model to the input layer).

In another study [**7**], the performances of seven well known machine learning methods in the tourism prediction problem are investigated. Furthermore, they investigate the effect of including the time index as an input variable.

During the last two decades, ANN and support vector regression (SVR) techniques have been started to immerge in the field of the tourism demand forecasting **[**8**,**9**,**10**,**8**,**1**,**11**,**12]. Because of their self-adaptive and nonlinear characteristic, machine learning techniques are becoming significant alternative tools for the seasonal time series forecasting against the traditional statistical methods. While some researchers have been suggesting that removing the seasonal patterns will increase the performance of the machine learning models [4,13], some others [14,6] are concluded that they can effectively deal with those patterns without seasonal adjustment of the raw data [3].

The most commonly used time series models for forecasting are traditional statistical methods. The drawback of these models is that they are mostly linear models. The relationship between the variables is not linear for most problems in real life [15] and using linear models for such problems is not efficient.

The statistical methods such as linear regression are suitable for data having seasonal or trend patterns, while artificial neural network (ANN) techniques are also efficient for data which are influenced by the special case, like promotion or extreme crisis [16].

The main purpose of this study is to deterministically generate the auxiliary variables in order to introduce the seasonal, cyclic and trend components of a seasonal time series to the multivariate machine learning models in the context of international tourism demand.

II. DATA and METHODOLOGY

In this study, dataset is composed of the monthly time series in a range of 1996 and 2013 years associated with Turkey and its top 10 ranked tourism clients. The number of tourists is selected as a metric to measure the tourism demand to Turkey.

Explanatory variables are: wholesale prices index, US Dollar selling, one ons gold London selling price in USD, hotel bed capacity of turkey, number of tourism agency in Turkey, harmonic consumer price index (HCPI) of Turkey and HCPIs of leading clients (namely France, Italy, Netherlands, United Kingdom, United States, number of the tourists coming from the top 10 leading clients of Turkey (namely Germany, Russia, England, Iran, Bulgaria, Georgia, Holland, France, USA, Italy), exchange rate of the leading countries of Turkey (British Pound, Russia Rouble, Bulgarian Lev), exchange rate of Turkish Lira, and the number of former tourists coming.

Monthly time series data were collected from the Ministry of the Tourism of the Republic of Turkey (website: www.turizm.gov.tr), State Institute of Statistics of Turkey (website: www.die.gov.tr), Databank of the Central Bank of the Republic of Turkey (website: http://evds.tcmb.gov.tr), TÜRSAB (website: www.tursab.org.tr).

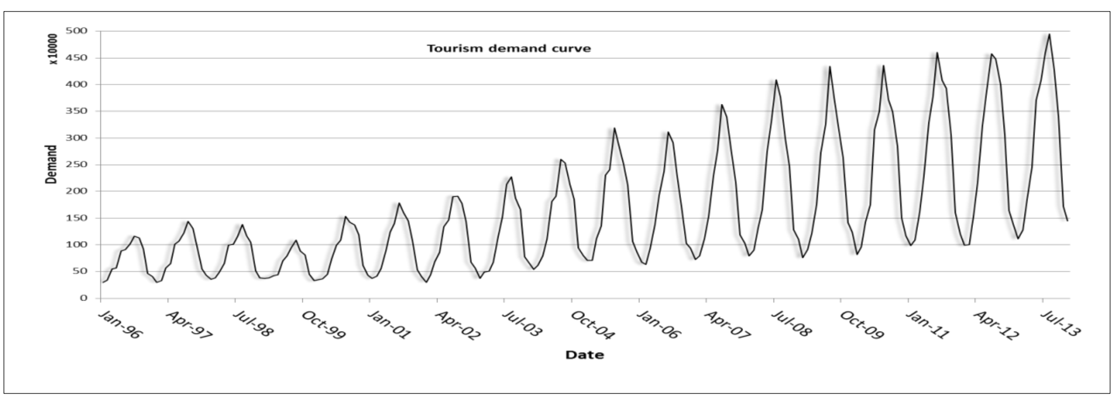

Fig. 1. Graphical representation of seasonal tourism demand data.

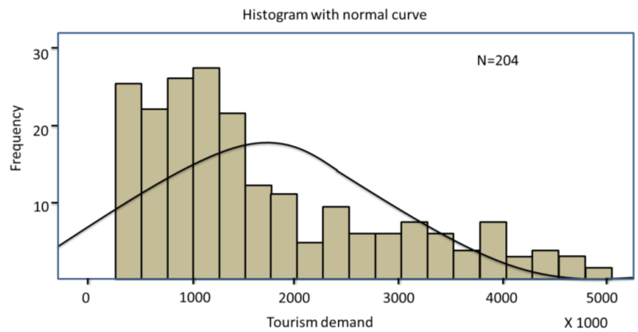

Fig. 2. Histogram of tourism demand data with normal curve.

Next to these real 29 time series including the target, deterministically four more auxiliary variables called year index (labelled A_cycle), month index (labelled M_cycle), annual average (labelled A_AVG) and monthly average (labelled M_AVG) are generated, which will be explained in the next section in details.

A. Analysis of Tourism Demand Data

The runs test used to detect whether the order of occurrence of two values of a variable is random. In runs test, a run represents a sequence of identical cases. If we apply this test to the continuous data, it also can indicate the existence of the trend in the data. We apply one-sample runs test on tourism demand data by setting the cut point to the median value of the tourism demand data series, so that values less than the mean would be in one group and the others in the second group. Number of the cases in each group is observed as 102. We see totally 204 observations with 30 runs. A sample with too many or too few runs indicates that the sample is not random. Here number of the runs (30), which is similar occurrence of the observations based on the cut point, is smaller than the number of each group (102). But since this is a large sample case, we consider the results of the test statistic (Z value = -9.967) and two tailed p value = 0.001 (significance level of 0.05) rather than the number of the runs. We can conclude that tourism demand data series has very strong evidence for trend.

The Kolmogorov-Smirnov one-sample test is one of the widely used tools to test seasonality in any annually cyclic phenomena. We have used The Kolmogorov-Smirnov one-sample test to detect the presence of any significant seasonal component in tourism demand time series based on the yearly period. In the one sample Kolmogorov-Smirnov Test, the "Asymp. Sig." (2-tailed) (also called as the p - value) is observed as 0.001 which is smaller than the significance level of 0.05, there is insufficient evidence to suggest the tourism demand series does not exhibit seasonal patterns.

In addition to the formal tests, we can also examine data graphically for trend, seasonal and cyclic components. From the graphical representation of seasonal tourism demand data (Fig. 1), the curve of the histogram's shape (Fig. 2) and monthly seasonal subseries plot (Fig. 3) [17], it is seen that there are strong evidences for existing of the trend and seasonal components in the tourism demand time series.

One of the most convenient ways is subseries plots with 12 categories representing the months on the x-axis and the usage of lines to display monthly tourism demand patterns as seen in Fig. 3.

B. Generation of the Auxiliary variables

We have comprehensively analysed the input data regards to the seasonal, trend and cyclic components. Since we have employed the multivariate data set, these components had been traced based on the tourism demand time series, which is associated with the time series of the total number of tourist coming to Turkey. The time series data for the total number of tourists coming to Turkey for the period from January 1996 to December 2013 exhibits very clear seasonality, cyclical effect and growth trend, as shown in Fig.s 1, 2, and 3 and Table 1. We have regarded these patterns as deterministic components and added them as auxiliary variables to the raw data set.

In the first step, we have computed the annual and monthly averages of tourism demand time series associated with time series of the total number of tourist coming to Turkey. To justify the dataset respect to the annual trend and monthly seasonal effects, we have added these averages to dataset by using two new variables called A_AVG (Annual average) and M_AVG (Monthly average). Generation of these variables is

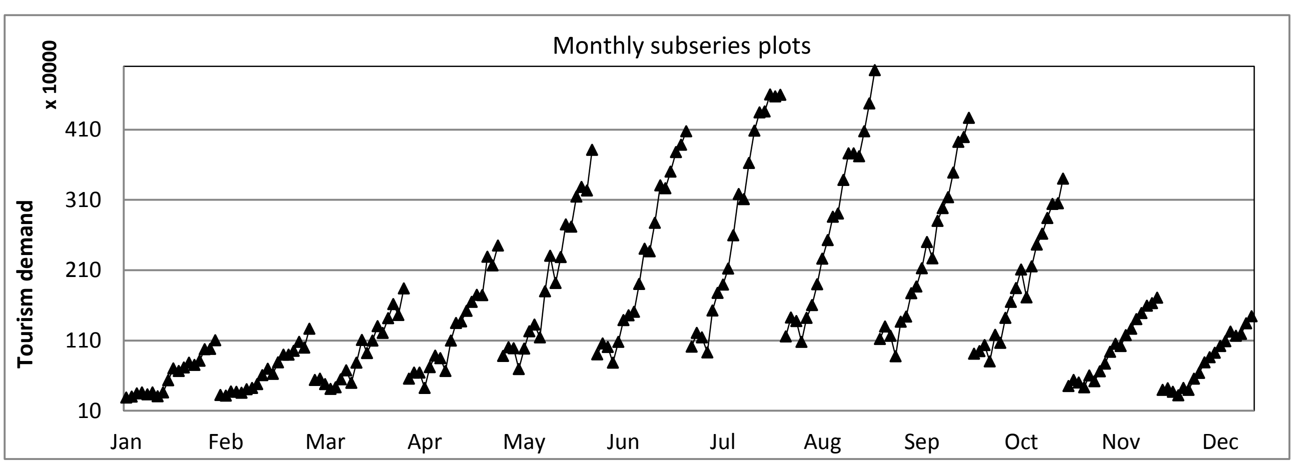

Fig. 3. Monthly subseries plots of tourism demand.

shown in the Table 1 and employments of them are illustrated in Table 2.

TABLE I
TOURISM DEMAND TIME SERIES WITH ANNUAL AND

Date	Jan	Feb	...	Nov	Dec	Annual Average
1996	288901	326411	...	453461	395022	711572
1997	302452	315746	...	540201	419123	810446
1998	348164	372963	...	501644	371636	785940
	⋮	⋮	⋮	⋮	⋮	⋮
2011	975723	1079505	...	1596295	1194729	2621340
2012	981611	997571	...	1631647	1343220	2648569
2013						
Monthly Average	563453	623233	...	945049	772983	1527228

In the second step, to introduce the cyclic patterns which are exhibited as yearly and monthly by the time series of tourism demand (labelled Y in the table 2), we have added two variables labelled A_cycle (Annual cycle) and M_cycle (Monthly cycle). These variables contain the year and month information respectively. These variables are illustrated in Table 2. One of the similar studies is [**7**], which investigates the effect of including the time index as an input variable.

C. Artificial Neural Networks Approach

ANNs are computing structures inspired from the biological neural networks. ANN is made of the interconnected processing units (usually called neurons). They have the ability of learning by adjusting the strength of the interconnections which can be achieved by altering the values called weights through the input data [18].

ANN is constructed with neurons. Each neuron has one or more weighted inputs (dendrites) and one or more outputs (axons) that are weighted when connecting to other neurons. Processing unit (Neuron) sums the weighted inputs and conveys the net input through an activation function in order to normalize and produce a result [19]. The equation of a simple neuron is given as

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

The multilayer network architecture consists of an input layer, two or more hidden layers, and one output layer. While the sigmoid function is used for the inner nodes, *linear activation function is used in output layer*. BP is one of the most popular approximation approache for training the multilayer feedforward neural networks based on the Widrow–Hoff training rule **[**20,18]. While the input signals propagate forward, error signals propagate backward layer-by-layer through the network [21].

There are two types of error functions for BP. The first error function (Eq. 2) is used for output cells, and the second is used only for hidden cells (Eq. 3).

$$E_0 = (a_i - y_i)\, g'(y_i) \tag{2}$$

$$E_h = \left(\sum_{i=1}^{i<n} \left(w_{h,i} E_0\right)\right) g(y_h) \tag{3}$$

where *"y"* is the output of the given cell. *"a"* is the expected or correct result. *"w"* represents all the weights (from 1 to n) connecting the hidden cell to all input cells (in a fully connected network). While *(g)* is the activation, or transfer function, function g' represents the first derivative of the activation function. The next step is to adjust the corresponding weights for the node by using this error. We'll use Eq. 4 for this purpose, which utilizes the error previously calculated for the node (whether hidden or output) [18].

$$w_{ij} = w_{ij} + \mu E y_i \tag{4}$$

For the given error *(E)* and activation (or cell output, y_i), we multiply by a learning rate (μ) and add this to the current

TABLE II
DATASET 2

Inputs→	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	...	X_{32}	Target→	Y
Date	A_cycle	M_cycle	A_AVG	M_AVG	HCPI_BG	HCPI_D	HCPI_F	HCPI_I	...	P_month	Date	T_Demand
Jan-96	1996	1	711572	563453	10.3	87.9	85.7	80.6	...	380022	Jan-97	302452
Feb-96	1996	2	711572	623233	10.3	88.4	86	81	...	288901	Feb-97	315746
Mar-96	1996	3	711572	885680	10.3	88.5	86.7	81.2	...	326411	Mar-97	557148
Apr-96	1996	4	711572	1195975	10.3	88.4	86.7	81.5	...	539738	Apr-97	643169
May-96	1996	5	711572	1864091	10.3	88.6	86.9	81.9	...	557846	May-97	1003366
Jun-96	1996	6	711572	2138993	10.3	88.7	86.8	82	...	878286	Jun-97	1059153
Jul-96	1996	7	711572	2711085	10.3	88.9	86.7	81.9	...	902258	Jul-97	1209037
Aug-96	1996	8	711572	2570000	10.3	88.8	86.5	82	...	1011137	Aug-97	1427984
Sep-96	1996	9	711572	2242218	10.3	88.8	86.7	82.1	...	1155810	Sep-97	1298700
Oct-96	1996	10	711572	1813971	10.3	88.7	87	82.2	...	1118972	Oct-97	949270
Nov-96	1996	11	711572	945049	10.3	88.6	86.9	82.5	...	911022	Nov-97	540201
Dec-96	1996	12	711572	772983	10.3	89	87.1	82.6	...	453461	Dec-97	419123
Jan-97	1997	1	810446	563453	15.3	89.5	87.3	82.8	...	395022	Jan-98	348164
Feb-97	1997	2	810446	623233	52.5	89.9	87.5	82.8	...	302452	Feb-98	372963
Mar-97	1997	3	810446	885680	56.5	89.8	87.6	83	...	315746	Mar-98	478063
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮		⋮
Nov-12	2012	11	2648569	945049	145.7	113.8	114	118.9	...	3050981	**Nov-13**	1709479
Dec-12	2012	12	2648569	772983	146.1	114.8	114.4	119.3	...	1631647	**Dec-13**	1442995

weight. The result is a minimization of the error at this cell, while moving the output cell activation closer to the expected output [19].

D. Support vector regression

SVMs are a new type of supervised learning methodology developed by Vapnik and his co-workers [22]. It is currently the most popular learning algorithm for supervised learning. SVMs build a linear separating hyperplane, by mapping data into the feature space with the higher-dimensional, using the so-called kernel trick. Even samples are not linearly separable in the original input space they can easily separate in the higher-dimensional space by using a linear separator. The high-dimensional linear separator is actually nonlinear in the original space [23].

Suppose we are given training data $\{(x_1, y_1),\ldots,(x_i, y_i)\} \subset \aleph x \Re$, where $\aleph$ denotes the space of the input patterns. In $\varepsilon - SV$ regression [22], the goal is to find a function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible. In other words, we do not care about errors as long as they are less than ε, but will not accept any deviation larger than this. The case of linear function $f(x)$ has been described in the form as is

$$f(x) = \langle w, x\rangle + b, \text{with} \quad \omega \in \aleph, \quad b \in \Re \tag{5}$$

where $\langle \cdot,\cdot \rangle$ denotes the dot product in $\aleph$. $\langle w, x\rangle$ is called feature, which is nonlinear mapped from the input space *x*. The *w* and *b* are coefficients, which are estimated by minimizing the regularized risk function. Flatness in the case of Eq. 8 means that one seeks a small *w*. The learning procedure of a SVM can be shown as follows. The minimization of the complexity term can be achieved by minimizing the quantity

$$\frac{1}{2}\|\overline{w}\|^2 \tag{6}$$

One way to ensure this optimizing by the way:

$$\text{minimize} \qquad \frac{1}{2}\|w\|^2$$

$$\text{subject to} \begin{cases} y_i - \langle w, x_i\rangle - b \le \varepsilon \\ \langle w, x_i\rangle + b - y_i \le \varepsilon \end{cases} \tag{7}$$

The *w* and *b* are coefficients, which are estimated by minimizing the regularized risk function

$$R(C) = \left(\frac{C}{N}\right)\sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \frac{\|w\|^2}{2}, \tag{8}$$

where,

$$L_\varepsilon(d_i, y_i) = \begin{cases} 0 & |d-y| \le \varepsilon \\ |d-y| - \varepsilon & \textit{otherwise}, \end{cases} \tag{9}$$

$L_\varepsilon(d_i, y_i)$ is called the e-insensitive loss function. *C* and *ε* are user-defined parameters. In the empirical analysis, *C* and *ε* are the parameters selected by users. The parameter *ε* is the difference between actual values and values calculated from the regression function. This difference can be treated as a tube around the regression function. The points outside the tube are the training errors. The loss equals zero if the forecasted value is within the e-tube **[**24,25**].**

E. Prediction Performance Metrics

There are a number of possible measures used for comparing the performances of MLP regression, and SVR models: root relative squared error (RRSE), relative absolute error (RAE), and correlation coefficient (*R*, sometimes also denoted *r*), respectively. These measures are defined in the following formulas, where the n is the number of test cases, a_i is the actual (observed) value, p_i is the predicted (estimated) value for the test case *i* [23]:

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

$$S_{\mathrm{A}} = \frac{1}{n-1}\sum_{i=1}^{n}(a_i - \bar{a})^2, \quad S_{\mathrm{P}} = \frac{1}{n-1}\sum_{i=1}^{n}(p_i - \bar{p})^2, \tag{10}$$

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

Root relative squared error (RRSE)

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

Relative absolute error (RAE)

$$RAE = \frac{\sum_{i=1}^{n}|p_i - a_i|}{\sum_{i=1}^{n}|a_i - \bar{a}|} \tag{12}$$

Correlation coefficient (R)

$$R = \frac{S_{PA}}{S_P S_A} \tag{13}$$

III. RESULTS and DISCUSSION

In this study, we have developed the MLP regression and SVR models to forecast the tourism demand to Turkey with respect to its ten major clients by using a wide variety of time series for a period concerning the years between 1996 and 2013 with monthly collected historical data.

In the experiments, the several different configurations of the MLP regression and SVR models are implemented. Their forecasting performances are examined and compared. In our implementations, we have used WEKA [26] data mining software, which is open source software issued under the GNU General Public License.

A. Experimental Results

In this study, ANN and SVR models are developed and tested on the tourism data set with 28 features for the forecast horizons of 12 months ahead. All the 28 time series are accumulated from the statistical institutions. The selection of the input variables in the development of the model significantly influences the model performance. “Regressional ReliefF” [27] data mining analysis was employed to determine the set of the input variables. Thus, the number of the tourists coming from the top ten client countries, exchange rates, harmonic consumer price indexes, wholesale prices index, gold price, hotel bed capacity, number of tourism agencies in destination place were consistently used as the input variables for training the machine learning techniques throughout the modelling phase.

Arrangement of the first dataset is composed of the 28 real time series, which are gathered from the statistical institutions, and the pairs of corresponding output with 12 months horizon. Arrangement of the second dataset is composed of the 28 real time series, which are the same variables in the first dataset, and four more auxiliary variables, which are deterministically generated and their pairs of corresponding output with 12 months horizon (Table 2). Instances of these datasets are respectively defined by

[x(t,1), x(t,2), … , x(t, 28); y(t+12)] and [x(t,1), x(t,2), … , x(t, 32); y(t+12)].

On the basis of MLP regression and SVR models and selection of their corresponding parameters, several models are implemented and examined. But because of space restriction, only the best of them in those combinations are reported in Table 3. Those models were evaluated with 10-folds cross validation by using three forecasting accuracy measures: correlation coefficient (R), relative absolute error (RAE) and root relative squared error (RRSE).

The MLP regression models developed in this study are constructed as two hidden layers with the sigmoid activation function and an output layer with a linear activation function. The first MLP regression model is implemented with the configuration of 28 (which is the number of the input variables in the first dataset) and18 nodes in the hidden layers, and trained by using back-propagation algorithm with the settings of the learning rate 0.1, momentum 0.7, and epoch 500. The second MLP regression model is implemented with the configuration of 32 nodes (which is the number of the input variables in the second dataset) and 14 nodes in the hidden layers, and trained by using back-propagation algorithm with the settings of the learning rate 0.1, momentum 0.7, and epoch 500.

The first SVR model is implemented by using the PUK kernel and with the settings of the complexity parameter c=150, omega parameter $\omega = 7$ and sigma parameter $\sigma = 14$ and the second SVR model is implemented by using the PUK kernel and with the settings of the complexity parameter c=150, omega parameter $\omega = 7$ and sigma parameter $\sigma = 14$. Forecasting results of the MLP and SVR models are given in Table 3 by means of relative absolute error (RAE), root relative squared error (RRSE) and the correlation coefficient (R).

Investigation and evaluation of those models (Table 3) showed that (1) ANN model using the second dataset with R=0.9879, RAE=14.41% and RRSE=16.27% values has better accuracy than ANN model using the first dataset with R=0.9821, RAE=16.96% and RRSE=19.11%, (2) SVR model using the second dataset with R=0.9932, RAE=10.36% and RRSE=11.66% values has better accuracy than SVR model using the first dataset with R=0.9875, RAE=14.57% and RRSE=15.84%, (3) the SVR model using the second dataset has the best accuracy with R=0.9932, RAE=10.36% and RRSE=11.66% values.

Like prior studies, which are concluded that machine learning models like ANN and SVR can effectively deal with the datasets having seasonal, trend and cycler patterns [3] and those techniques are also efficient for data, which are influenced by the special case, like promotion or extreme crisis [16], this study is also showed that natural performance of MLP regression and SVR models appears to be satisfactory (which can be seen in table 3, model 1 and 3). However, for the better forecasting performance, the further investigation has done based on the trend, seasonal, and cyclic components of the tourism demand time series (which can be seen in table 3, model 2 and 4).

We have enriched the raw data set, which is called dataset I and contains 28 real data series by adding deterministically generated four more auxiliary variables by considering the date information and patterns involved in tourism demand time series and we have obtained a new data set, which is called dataset II with 28 real attributes and four artificial attributes. The performances of the ANN and SVR models are evaluated for each datasets.

Choosing the large size of dataset may require the consideration of the trade-offs between the speed and accuracy. Larger dataset takes longer to train and to generate forecast. If the consideration is not only accuracy of the forecasting model, the smaller number of input variables could be preferred. In the case of more accuracy, losing some performance, which is a few seconds for this study, can be neglected.

The purpose of this study was to perform a comprehensive analysis for a tourism demand time series to investigate whether it has trend, seasonal and cyclic components. And then we have developed deterministic techniques in order to outline and introduce these patterns to the machine learning models.

TABLE III
FORECASTING PERFORMANCES

Model No	Dataset	Model Type	Configuration	R	RAE (%)	RRSE %
1	Dataset I	SVR	PUK kernel, C:150, O/S:7/14	0.9875	14.57	15.84
2	Dataset II	SVR	PUK kernel, C:150, O/S:7/14	0.9932	10.36	11.66
3	Dataset I	ANN	ANN(28,18), M:0.7, LR:0.1	0.9821	16.96	19.11
4	Dataset II	ANN	ANN(32,14), M:0.7, LR:0.1	0.9879	14.41	16.27

We have investigated the applicability of these techniques in machine learning area by employing the ANN and SVR models, which are well-known machine learning models in the tourism demand forecasting.

IV. CONCLUSIONS

We have analyzed the tourism demand series by considering the certain possible patterns. The purpose of analyzing the time series is to introduce the dynamic relationships among the time and the tourism demand series and improve the accuracy of forecasts by extracting the additional information available from the associated data series at hand in the tourism demand forecasting.

In tourism demand forecasting, the quantitative forecasting models are mainly studied and examined in three general categories: time-series, econometric techniques, and artificial intelligence (AI) methods. The seasonal, cyclic and trend components of the seasonal time series are intensively investigated and considered in tourism demand modelling and forecasting for time-series and econometric models. To deal with these patterns, many techniques such as lagged explanatory variables, dummy variables, seasonal indexes, deseasonalization and differencing are developed. However, little empirical research has been undertaken on incorporating of these patterns when the machine learning models are used in the modelling of the tourism demand. This empirical study uses the data series with monthly data points, and it tests the appropriateness of the two proposed techniques. These proposed techniques aim to outline the trend, seasonal and cyclic components and introduce them to the machine learning models in the contents of tourism demand modelling and forecasting.

The conclusion drawn from this study is that two proposed techniques, which are inclusion of two date dimensions regards to the yearly and monthly cycles, and two justifications of data regards to trend outlined by the annual averages and seasonal pattern outlined by the monthly averages, significantly improve the accuracy of the machine learning models in the context of the tourism demand forecasting. This outstanding improvement is obtained based on a particular data set related to the tourism demand for Turkey. Therefore, this study can be extended on the other time series which exhibits the cyclic, seasonal and trend patterns in the use of machine learning models.

One of the deficiencies of using these techniques is a requirement of exploratory data analysis to extract information for cyclic, seasonal and trend components, which is time-consuming and requires domain experts. However, the main motivation of the developing machine learning algorithms is to introduce the tools that automate the information discovery process. But the significant improvement of the performance regards to accuracy obtained by the use of these methods shows that these and other similar statistical techniques are challenging pre-processing considerations in the development of the machine learning models.

REFERENCES

[1] Haiyan Songa and Gang Li, "Tourism demand modelling and forecasting - A review of recent research," Tourism Management, no. 29, pp. 203–220, 2008.
[2] Rob J Hyndman, "Forecasting overview," 2009.
[3] R Adhikari and RK Agrawal, "Forecasting strong seasonal time series with artificial neural networks," Journal of Scientific & Industrial Research, vol. 71, pp. 657-666, 2012.
[4] Tim Hill, Marcus O'Connor, and William Remus, "Neural Network Models for Time Series Forecasts," Management Science, vol. 42, no. 7, pp. 1082-1092, 1996.
[5] J.H. Wang and J.Y. Leu, "Stock market trend prediction using ARIMA-based neural networks," in IEEE Int. Conf. Neural Networks , 1996, pp. 2160–2165.
[6] Fang-Mei Tseng, Hsiao-Cheng Yu, and Gwo-Hsiung Tzeng, "Combining neural network model with seasonal time series ARIMA model," Technological Forecasting and Social Change, vol. 69, no. 1, pp. 71-87, 2002.
[7] Nesreen K. Ahmed, Amir F. Atiya, Neamat El Gayar, and Hisham El-Shishiny, "Tourism Demand Foreacsting Using Machine Learning Methods," International Journal on Artificial Intelligence and Machine Learning, pp. 1-7, 2008.
[8] R. Law and N. Au, "A Neural Network Model to Forecast Japanese Demand for Travel to Hong Kong," Tourism Management, vol. 20, pp. 89-97, 1999.
[9] R. Law, "Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting," Tourism Management, vol. 21, no. 4, pp. 331-340, 2000.
[10] D. Pattie and J. Snyder, "Using a Neural Network to Forecast Visitor Behavior. Annals of Tourism Research," 1996.
[11] P. F. Pai and W. C. Hong, "An improved neural network model in forecasting tourist arrivals," Annals of Tourism Research, 32, 1138–1141, 2005.
[12] Kuan-Yu Chen and Cheng-Hua Wang, "Support vector regression with genetic algorithms in forecasting tourism demand," Tourism Management, pp. Volume 28, Issue 1, February 2007, Pages 215–226, 2005.
[13] G. Peter Zhang and Min Qi, "Neural network forecasting for seasonal and trend time series," European Journal of Operational Research, vol. 160, no. 2, pp. 501-514, 2005.
[14] Ilan Alona, Min Qi, and Robert J. Sadowski, "Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods," Journal of Retailing and Consumer Services, vol. 8, no. 3, pp. 147–156, 2001.
[15] C.W.J. Granger and T. Terasvirta, Modelling Nonlinear Economic Relationships: Oxford University Press, 1993.
[16] Tugba Efendigil, Semih Önüt, and Cengiz Kahraman, "A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis," Expert Systems with Applications 36 (2009) 6697–6707, 2009.
[17] William Cleveland, Visualizing Data.: Hobart Press, 1993.

[18] S. Haykin, Neural Networks: a comprehensive foundation., Second Edition ed.: Prentice Hall, 1999, 842p.
[19] M. Tim Jones, Artificial Intelligence: A Systems Approach.: Infinity Science Press LLC, 2008.
[20] Christopher M. Bishop, Neural Networks for Pattern Recognition.: Oxford University Press, 1995.
[21] Atilla Aslanargun, Mammadagha Mammadov, Berna Yazici, and Senay Yolacan, "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, 2007.
[22] V. Vapnik, The nature of statistical learning theory. New York: Springer, 1995.
[23] Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed.: Morgan Kaufmann, 2005.
[24] Harris Drucker, Chris J.C. Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik, "Support vector regression machines," Adv. Neural Inform. Process. Syst. 9 (1997) 155–161., 1997.
[25] Vojislav Kecman, Learning and Soft Computing:Support Vector Machines, Neural Networks and Fuzzy Logic Models.: A Bradford Book 1 edition ISBN-10: 0262112558, 2001.
[26] Mark Hall et al., "The WEKA data mining software: an update," ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. Volume 11 Issue 1 ACM New York NY USA ISSN: 1931-0145, 2009.
[27] Marko Robnik-Šikonja and Igor Kononenko, "An adaptation of Relief for attribute estimation in regression," 1997.

BIOGRAPHIES

Selcuk CANKURT is graduated from the University of Marmara, Istanbul, Turkey in 1997. He received the M.S. degree in information technologies from International Burch University, Sarajevo, Bosnia and Herzegovina in 2011. Since 2013, he has appointed as senior teaching assistant in International Burch University. His research interest is multivariate forecasting in the areas of machine learning, data mining and soft computing.

Abdulhamit SUBASI received the B.S. degree in electrical and electronics engineering from Hacettepe University, Ankara, Turkey, in 1990 and the M.S. degree in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 1993. He received his Ph.D. degree in electrical and electronics engineering from Sakarya University, Sakarya, Turkey, in 2001. In 2006, he was senior researcher at Georgia Institute of Technology, School of Electrical and Computer Engineering, USA. Since 2012, he is appointed as Professor of Electrical and Electronics Engineering at International Burch University, Sarajevo, Bosnia and Herzegovina. His areas of interest are data mining, machine learning, pattern recognition and biomedical signal processing. He has worked on several projects related with biomedical signal processing, pattern recognition and computer network security. Moreover, he is voluntarily serving as a technical publication reviewer for many respected scientific journals and conferences.