An Application of Artificial Neural Networks for Prediction and Comparison with Statistical Methods

S. Ballı[1], İ. Tarımer[1]

[1]Department of Information Systems Engineering, Faculty of Technology, Muğla Sıtkı Koçman University,
48170, Muğla, Turkey, phone: +90 252 211 31 06
serkan@mu.edu.tr

Abstract—The unemployment rate is a measure of the prevalence of unemployment and can be a good indication of the for country's economic situation. It is thought that it might change depending upon educational situations of the people. This study seeks how to model and predict the unemployment rates with respect to educational situations of the people in Turkey. For this purpose, An Artificial Neural Networks model was proposed for prediction long-term prediction (up to year 2019) is performed by using Artificial Neural Networks, Box-Jenkins Method and Regression Analysis Method. Then these methods have been compared with each other; and the study concludes that Artificial Neural Networks is more appropriate and consistent than Box-Jenkins Method or Regression Analysis Method for the prediction.

Index Terms—Artificial neural networks, prediction, Box-Jenkins method, regression analysis method.

I. INTRODUCTION

Unemployment and recruitment are the key indicators of economic development level for national economies. In many countries including Turkey, unemployment is considered as a major source of social problems and unemployment problem must be solved immediately. A productive economy and greater competitiveness are referred to as the number one condition for becoming a developed country. Countries suffering from unemployment must improve human resources for economic development [1]. Prediction of the unemployment rates is very important utility for the countries.

Statistical methods such as Box-Jenkins Method (BJM) [2] and Regression Analysis Method (RAM) can be used for the prediction. However, according to previous researches on prediction and modelling of non-linear systems, BJM and RAM are discovered to be insufficient [3]. Artificial intelligence techniques are convenient to overcome the disadvantages of BJM and RAM [4], [5] and viewed as the most promising methods towards intelligent systems [6]. Artificial Neural Networks (ANN) is one of those approaches implemented via computer technology in all

fields; and this is actually preferred to other methods. ANN is used for the prediction in various applications. Seyhan et al. [7] employed ANN model to predict the effects of the thermoplastic binder concentration. Karunasinghea and Liong [8] made chaotic time series prediction with artificial neural networks. Bezerra et al. [9] predicted kinetic parameters of carbon reinforced fiber composites by using ANN. Karazi et al. [10] used ANN approach for the prediction of laser-machined micro-channel dimensions. Brown and Moshiri [11] applied two ANN models, a back-propagation model and a generalized regression neural network model to estimate and forecast unemployment rates. Ioana and Atsalakis [12] presented an ANN and fuzzy inference system to forecast of Greek unemployment rate. Wang and Zheng [13] used back-propagation neural network to predict unemployment rate of China.

This study investigates how unemployment rate changes with respect to the people's educational situations by using artificial neural networks. To achieve the projected end, data sets, between the years 1992 and 2008 were taken from Turkish Statistical Institute [14]. Then an ANN model was proposed for prediction and BJM, RAM and ANN methods were respectively applied for the prediction of unemployment rates with respect to educational situations of the people.

II. ARTIFICIAL NEURAL NETWORKS

ANN is designed in a similar format to that of human brain; they are systems consisting of a number of processing units and operating parallel. The main principle of ANN is based on finding coefficients between the inputs and outputs of a problem, making connections between input and output layers and doing all jobs on a learning system [15]. The ANN has the ability to establish the relationship between input-output data that identified depending on the parameters of a system [16].

One of the advantages of ANN, it does not need detailed information about the system. ANN can also define the complicated and complex relations inside the system. Recently, ANN has been applied to a number of different problems that involves transactions on classifying, predicting, control systems, optimization and decision

making [1], [17]. Another advantage of ANN is that it is independent of statistical distribution of the data [18].
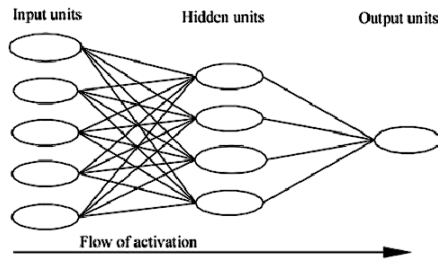

Fig. 1. Basic ANN model.

ANN basically consists of three layers as shown in Fig. 1. The first layer consists of processing units related to input variables (cell or neuron); this section is called input layer. The function of input layer is to transmit the input variables to hidden layer coming after itself in the network. The final layer consists of output variables called the output layer. The layer consisting of the processing units between the input layer and the output layer is called the hidden layer. The presence of the hidden layer is useful in the modelling of the complex relations.

The number of the neurons in the layers varies depending on the complexity of the problem. Links coming out of each neuron of the input layer have weights; and the weights connecting z_h hidden layer with x_j input layer are labelled as w_{ij}. Each hidden layer neuron calculates the weighted sum of the neurons in the input layer; and this is given in (1)

$$Input_j^s = \sum w_{kj} x_k^i. \qquad (1)$$

Outputs corresponding to these layers are obtained as a result of the implementation of the function known as activation or transfer function to inputs. An ANN with randomly set weights is "dull" but can be trained by successive repetitions of the same problem. A network "learns" by iteratively correcting the weights (the only adjustable parameters) so as to produce the previously specified output values (target sets) for as many input sets as possible.

If the difference between these values designed as output and actual values is at the desired level, the algorithm is terminated; otherwise, the weights are updated in such a way that the default between these values is minimized. This algorithm is called back propagation; and this is the most commonly used learning algorithm in ANN [15], [17], [19].

A commonly used equation for calculating the error of a network is given in (2)

$$E = \frac{1}{2} \sum_j (t_j - Output_j)^2, \qquad (2)$$

where t_j and $output_j$ are the actual and desired values of unit j in the output layer. The weights are updated according to the so-called delta-rule of learning

$$\Delta w_{ij} = \eta \delta_i o_i, \qquad (3)$$

where $\eta > 0$ is the learning rate, δ_i is a correction term, and o_j is the output of unit i in the previous layer. The value of δ_i should be proportional to the output-error. In the back-propagation algorithm the correction-term is obtained by applying the so-called gradient descent method, which leads to the following expression for the delta-term of an output unit

$$\delta_i = (\partial E / \partial o_j)(\partial o_j / \partial I_j) = (t_j - o_j^{out}) o_j (1 - o_j). \qquad (4)$$

The following recursive formula is applied to calculate the correction term for a hidden unit

$$\delta_j = o_j (1 - o_j) \sum_k \delta_k w_{kj}. \qquad (5)$$

It has been found that the performance and also the stability of a training process are greatly enhanced if a so-called momentum term is added to the learning rule

$$\Delta w_{ij} = \eta \delta_i o_i + \mu \Delta w_{ij}^{prev}, \qquad (6)$$

where $0 < \mu < 1$ is a constant called momentum, and Δw_{ij}^{prev} is the adjustment to the same weight in the previous iteration cycle [15], [17], [19].

III. PREDICTION OF THE INVESTIGATED DATA

In this study, the rates of unemployment during the period of 2002-2008 are analyzed in connection with the people's educational situations. To this end, in order to predict the rates of unemployment ANN, BJM and RAM are used and the results are compared. For each method, the data set for the period from 1992 to 2002 is used as training set and the data set for the period from 2002 to 2008 is used as test set. For the analysis, MATLAB, EVIEWS and S-PLUS software programs are used.

For each educational situation, different models are used and each model is formed as follows:

1) *Model 1.* The year is chosen as the input variable; and the output variable is unemployment rate for the primary school graduates considered.
2) *Model 2.* The year is chosen as the input variable; and the output variable is unemployment rate for the secondary (high school) graduates.
3) *Model 3.* The year is chosen as the input variable; and the output variable is unemployment rate for the university graduates.

A. Results for ANN

The structure of the ANN proposed for the models is as follows: The neuron number of the input layer is 1, the number of hidden layer is 1, for model 1 and 2 the number of the neuron is 2 and for model 3, and it is 3 whereas for all the models the number of output layer neuron is 1.

Activation function used between the input layer and hidden layer is hyperbolic tangent and the activation function used between hidden layer and output layer is a linear function as shown in Fig. 2. Learning rate and

momentum rate are taken as 0.5; and as a learning algorithm Batch Gradient Descent algorithm is used. This operation is completed in 1000 iteration steps. Estimated values obtained through ANN method for test data set (2005-2008) for Model 1-3 are shown in Table I.
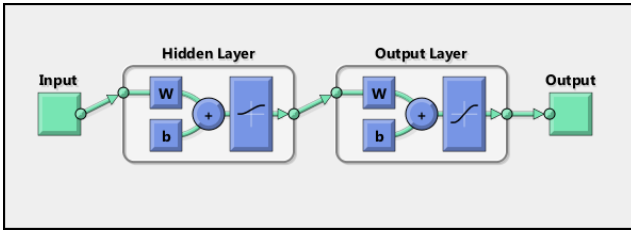

Fig. 2. Structure of ANN model.

TABLE I. ACTUAL AND ESTIMATED VALUES FOR MODEL 1–3.

Year	Model 1		Model 2		Model 3	
	Actual	Estimated	Actual	Estimated	Actual	Estimated
2005	0,079	0,102	0,192	0,191	0,077	0,097
2006	0,098	0,096	0,188	0,200	0,111	0,110
2007	0,102	0,093	0,163	0,190	0,111	0,078
2008	0,089	0,092	0,209	0,178	0,124	0,102

In order to compare the obtained results, Absolute Deviation (AD) and Mean Absolute Deviation (MAD) are used. AD and MAD values are given in Table II.

TABLE II. AD AND MAD VALUES.

Years	Model 1	Model 2	Model 3
	AD	AD	AD
2005	0,023	0,001	0,020
2006	0,001	0,011	0,001
2007	0,009	0,027	0,032
2008	0,003	0,031	0,022
MAD	0.009	0.018	0,019

Different topologies are used for these models; and they are compared through Mean Square Error (MSE). The obtained results for Model 1-3 are presented in Table III.

TABLE III. DIFFERENT TOPOLOGIES FOR MODEL 1-3.

		Topology		
		1-2-1	1-3-1	1-4-1
Model 1	Learning Rate	0,5	0,4	0,3
	Momentum	0,5	0,4	0,3
	MSE	$1,5054 \times 10^{-4}$	$1,5997 \times 10^{-4}$	$1,6394 \times 10^{-4}$
Model 2	Learning Rate	0,5	0,4	0,3
	Momentum	0,5	0,4	0,3
	MSE	$4,8583 \times 10^{-4}$	$6,9122 \times 10^{-4}$	$6,9807 \times 10^{-4}$
Model 3	Learning Rate	0,4	0,5	0,3
	Momentum	0,4	0,5	0,3
	MSE	$6,9948 \times 10^{-4}$	$4,821 \times 10^{-4}$	12×10^{-4}

Finally, long-term estimation of the unemployment rates in line with the levels of education over the period between 2009 and 2019 is calculated; and these results are shown in Table IV and presented graphically in Fig. 3. It's seen that the rates are fluctuating for Model 2-3 and stable for Model 1.

TABLE IV. THE RESULTS OF LONG-TERM ESTIMATION FOR 2009-2019.

Years	Model 1	Model 2	Model 3
2009	0,098	0,140	0,109
2010	0,099	0,141	0,109

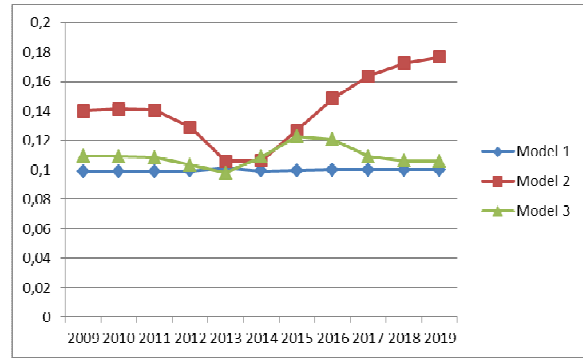| Years	Model 1	Model 2	Model 3
2011	0,099	0,140	0,108
2012	0,099	0,129	0,103
2013	0,101	0,106	0,098
2014	0,099	0,106	0,109
2015	0,100	0,127	0,123
2016	0,100	0,148	0,121
2017	0,100	0,163	0,109
2018	0,100	0,172	0,106
2019	0,100	0,177	0,106

Fig. 3. Estimation for 2009-2019 by using ANN.

B. Results for RAM

Regression equations obtained for Model 1-3 in regression analysis technique respectively are given in (7)–(9):

$$Y_{Model\ 1} = 6.180 - 0.00307\ X, \tag{7}$$
$$Y_{Model\ 2} = 25.189 - 0.0125\ X, \tag{8}$$
$$Y_{Model\ 3} = -4.165 + 0.002119\ X. \tag{9}$$

The estimated values obtained for regression analysis for 2005-2008 are given in Table V.

TABLE V. ACTUAL AND ESTIMATED VALUES FOR MODEL 1-3.

Years	Model 1		Model 2		Model 3	
	Actual	Estimated	Actual	Estimated	Actual	Estimated
2005	0,079	0,037	0,193	0,176	0,077	0,075
2006	0,098	0,034	0,188	0,164	0,110	0,077
2007	0,102	0,031	0,163	0,151	0,110	0,079
2008	0,089	0,027	0,209	0,139	0,123	0,081

AD and MAD values related to the estimations of this technique are shown in Table VI.

TABLE VI. AD AND MAD VALUES

Years	Model 1	Model 2	Model 3
	AD	AD	AD
2005	0,042	0,016	0,002
2006	0,064	0,024	0,033
2007	0,071	0,011	0,031
2008	0,061	0,070	0,041
MAD	0,060	0,030	0,027

The results of long-term estimation till 2019 are as seen in Table VII and presented graphically in Fig. 4. It's seen that the rates are in declining trend for each model.

TABLE VII. THE RESULTS OF LONG-TERM ESTIMATION FOR 2009-2019.

Years	Model 1	Model 2	Model 3
2009	0,030	0,120	0,130
2010	0,030	0,100	0,110
2011	0,020	0,090	0,100
2012	0,020	0,080	0,090
2013	0,020	0,070	0,080

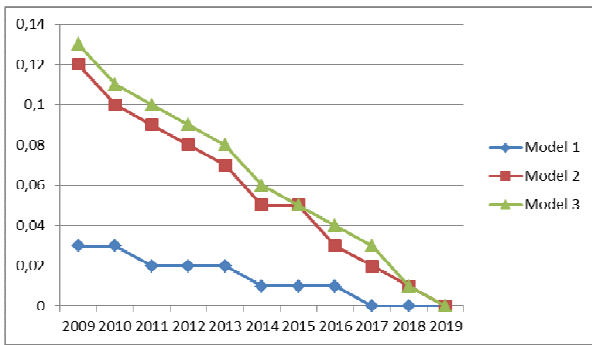| Years	Model 1	Model 2	Model 3
2014	0,010	0,050	0,060
2015	0,010	0,050	0,050
2016	0,010	0,030	0,040
2017	0,000	0,020	0,030
2018	0,000	0,010	0,010
2019	0,000	0,000	0,000

Fig. 4. Estimation for 2009-2019 by using RAM.

C. Results for BJM

BJM used for Model 1-3 out of the implementation of the relevant method are as follows:

When the first differences are taken for Model 1, it is concluded that the series tends to be stable and when its correlogram is drawn, there is only a significant observation in Autocorrelation Function graph (ACF), as shown in Fig. 5 and no significant observation in sinusoidal decreasing Partial Autocorrelation Function graph (PACF) as shown in Fig. 6. As the most suitable model for this, AR(0,1,1) and AR(0,2,1) models are considered the best model that fits. Since MAD is observed to be the smallest in AR(0,2,1) model, this model is used.

The initial differences are taken for Model 2; and it is observed that the series becomes stable. Moreover, based on its correlogram, it is concluded that there is one significant observation in ACF graph, and no significant observation in sinusoidal decreasing PACF graph. Therefore, AR(0,1,1) and AR(0,2,1) models are considered the best model fit. Since MAD is observed to be the smallest in AR(0,1,1) model, this model is used. After the first differences are taken for Model 3, it is observed that the series becomes stable. Based on its correlogram, it is concluded that there is one significant observation in ACF graph and no significant observation in sinusoidal decreasing PACF graph. As the most suitable model for this, AR(0,1,1) and AR(0,2,1) models are considered the best model fit. Since MAD is observed to be the smallest in AR(0,1,1) model, this model is used. The estimated values obtained by means of this technique for 2005-2008 are presented in Table VIII.

TABLE VIII. ACTUAL AND ESTIMATED VALUES FOR MODEL 1-3.

Years	Model 1 ARIMA(0,2,1)		Model 2 ARIMA(0,1,1)		Model 3 ARIMA(0,1,1)	
	Actual	Estimated	Actual	Estimated	Actual	Estimated
2005	0,080	0,064	0,193	0,122	0,078	0,122
2006	0,098	0,062	0,188	0,122	0,111	0,122
2007	0,102	0,061	0,163	0,122	0,111	0,122
2008	0,090	0,059	0,209	0,122	0,124	0,122

AD and MAD values obtained for this method are shown in Table IX. Finally, long-term estimation results for 2009-

2019 are presented in Table X and presented graphically in Fig. 7. It's seen that the rates are stable for Model 2-3 and declining for Model 1.
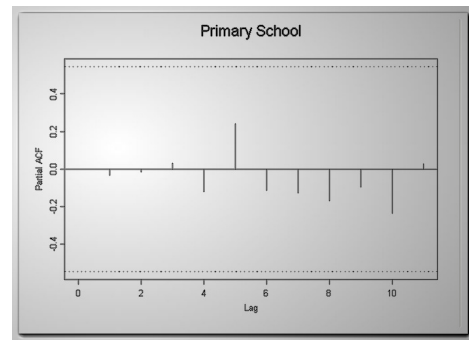

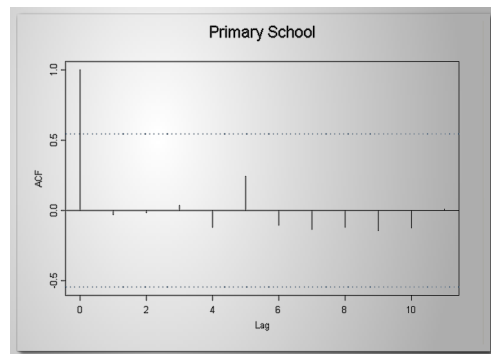Fig. 5. Autocorrelation function for Model.

Fig. 6. Partial autocorrelation function for Model 1.

TABLE IX. AD AND MAD VALUES.

Years	Model 1	Model 2	Model 3
	AD	AD	AD
2005	0,015	0,071	0,071
2006	0,035	0,066	0,066
2007	0,041	0,042	0,042
2008	0,030	0,087	0,087
MAD	0,030	0,067	0,067

TABLE X. THE RESULTS OF LONG-TERM ESTIMATION FOR 2009-2019.

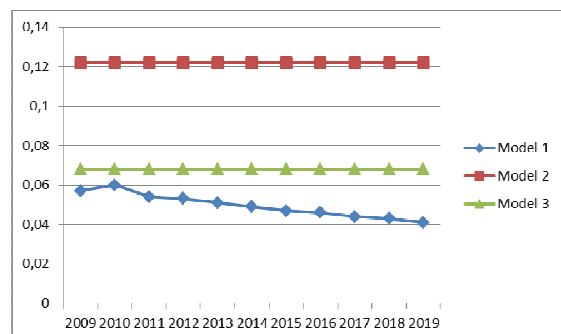| Years	Model 1	Model 2	Model 3
2009	0,057	0,122	0,068
2010	0,060	0,122	0,068
2011	0,054	0,122	0,068
2012	0,053	0,122	0,068
2013	0,051	0,122	0,068
2014	0,049	0,122	0,068
2015	0,047	0,122	0,068
2016	0,046	0,122	0,068
2017	0,044	0,122	0,068
2018	0,043	0,122	0,068
2019	0,041	0,122	0,068

Fig. 7. Estimation for 2009-2019 by using BJM.

D. Comparison of the methods

The minimum mean absolute error values for each method are given in Table XI and presented graphically in Fig. 8.

TABLE XI. MAD VALUES FOR THREE METHODS.

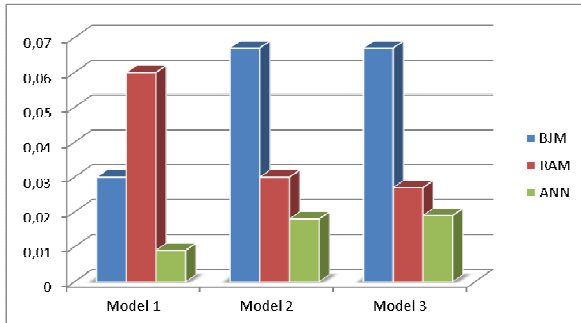| 2005-2008	BJM	RAM	ANN
Model 1	0,030	0,060	0,009
Model 2	0,067	0,030	0,018
Model 3	0,067	0,027	0,019

Fig. 8. Comparison of the methods for Model 1-3.

According to these results, it's seen that ANN has better error values than BJM and RAM for each model. Estimated values of ANN are closest to actual values. Finally it's shown that prediction by using ANN is more appropriate than other methods.

IV. CONCLUSIONS

This study investigates development of artificial neural networks for prediction and how unemployment rate changes with respect to the people's educational situations. In long-term prediction for Model 1 and 3, unemployment rates will be stationary but they will rise for Model 2 in future. For test data set, estimated values of ANN are closest to the actual values for Model 1-3. It is shown that ANN is quite robust since it can recognize the complex relations from among great number of variables better than other conventional statistical methods: BJM and RAM.

Disadvantages of BJM and RAM include with stationarity, necessity of providing some assumptions and not getting consistent results in which data sets are fairly small. On the contrary, ANN can work with small data set; and ANN is actually preferred to other methods because it has significant and specific attributes such as generalization, learning from data, working with unlimited number of variables and no necessity to pre-information about the problem.

REFERENCES

[1] S. Ballı, N. Güler, N. Yakar, "A Prediction of Unemployment Rates By Using Artificial Neural Networks via People's Educational Background", in *Proc. of the International Conference on Modelling and Simulation,* Konya, Turkey, 2006, pp. 947–951.

[2] G. E. P. Box, G. M. Jenkins, *Time Series Analysis: Forecasting and Control.* San Francisco, Holden Day Inc., 1976, p. 575.

[3] C. C. Chiu, C. T. Su, "A Novel Neural Network Model Using Box-Jenkins Technique and Response Surface Methodology to Predict Unemployment Rate", in *Proc. of the Tenth IEEE International Conference Tools with Artificial Intelligence,* Taipei, Taiwan, 1998, pp. 74–80.

[4] D. A. Dickey, W. A. Fuller, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica,* no. 49, pp. 1057–1072, 1981. [Online]. Available: http://dx.doi.org/10.2307/1912517

[5] W. Enders, *Applied Econometric Time Series.* Jonh Wiles and Sons, Canada, 2003, p. 480.

[6] A. G. Ak, G. Cansever, A. Delibasi, "Robot Trajectory Tracking With Adaptive RBFNN-Based Fuzzy Sliding Mode Control", *Information Technology And Control,* vol. 40, no. 4, pp. 151–156, 2011.

[7] A. T. Seyhan, G. Tayfur, M. Karakurt, M. Tanoğlu, "Artificial neural network (ANN) prediction of compressive strength of VARTM processed polymer composites", *Computational Material Science,* vol. 34, no. 1, pp. 99–105, 2005. [Online]. Available: http://dx.doi.org/10.1016/j.commatsci.2004.11.001

[8] D. S. K. Karunasinghea, S. Liong, "Chaotic time series prediction with a global model: Artificial neural network", *Journal of Hydrology,* vol. 323, no. 1–4, pp. 92–105, 2006. [Online]. Available: http://dx.doi.org/10.1016/j.jhydrol.2005.07.048

[9] E. M. Bezerra, M. S. Bento, J. A. F. F. Rocco, K. Iha, V. L. Lourenço, L. C. Pardini, "Artificial neural network (ANN) prediction of kinetic parameters of (CRFC) composites", *Computational Material Science,* vol. 44, no. 2, pp. 656–663, 2008. [Online]. Available: http://dx.doi.org/10.1016/j.commatsci.2008.05.002

[10] S. M. Karazi, A. Issa, D. Brabazon, "Comparison of ANN and DoE for the prediction of laser-machined micro-channel dimensions", *Optics and Lasers in Engineering,* vol. 47, no. 9, pp. 956–964, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.optlaseng.2009.04.009

[11] L. Brown, S. Moshiri, "Unemployment variation over the business cycles: a comparison of forecasting models", *Journal of Forecasting,* vol. 23, no. 7, pp. 497–511, 2004. [Online]. Available: http://dx.doi.org/10.1002/for.929

[12] U. C. Ioana, G. Atsalakis, "Soft Computing Methods Applied in Forecasting of Economic Indices Case Study: Forecasting of Greek Unemployment Rate Using an Artificial Neural Network with Fuzzy Inference System", in *Proc. of the 10th WSEAS International Conference on Mathematical and Computational Methods in Science and Engineering,* Bucharest, Romania, 2008, pp. 233–237.

[13] G. Wang, X. Zheng, "The Unemployment Rate Forecast Model Basing on BP Neural Network", in *Proc. of the International Conference on Electronic Computer Technology,* Macau, China, 2009, pp. 475–478.

[14] *TUIK, Statistical Indicators 1923-2008,* Turkish Statistical Institute, Ankara, 2009, p.720.

[15] A. Mellit, S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review", *Progress in Energy and Combustion Science,* vol. 34, no. 5, pp. 574–632, 2008. [Online]. Available: http://dx.doi.org/10.1016/j.pecs.2008.01.001

[16] B. Ciylan, "Determination of Output Parameters of Thermoelectric Module using Artificial Neural Networks", *Elektronika ir Elektrotechnika (Electronics and Electrical Engineering*), no.10. pp. 63–66, 2011.

[17] D. Graupe, *Principles of Artificial Neural Networks.* World Scientific Press, Singapore, 2007, p. 320.

[18] J. Daunoras, V. Gargasas, A. Knys, G. Narvydas, "Application of Neural Classifier for Automated Detection of Extraneous Water in Milk", *Elektronika ir Elektrotechnika (Electronics and Electrical Engineering*), no. 9, pp. 59–62, 2011.

[19] S. J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach.* Pearson Education Inc., New Jersey, 2003, p. 1132.