Turkish Journal of Forecasting vol. 05 no. 2 (2021) pp. 36-44

Content list available at JournalPark

Turkish Journal of Forecasting

Journal Homepage: tjforecast.xyz

Estimating CO_2 Emission Time Series with Support Vector Machines Regression, Artificial Neural Networks, and Classic Time Series Analysis

F. Cemrek[1,*], Ö. Demir[2]

[1]*Eskişehir Osmangazi University, Faculty of Science and Letters, Department of Statistics, Eskişehir, Turkey*
[2]*Eskişehir Osmangazi University, Graduate School of Natural and Applied Sciences, Department of Statistics, Eskişehir, Turkey*

A R T I C L E   I N F O

Article history:

Received	13	December	2021
Revision	27	December	2021
Accepted	28	December	2021
Available online	31	December	2021

Keywords:

Support vector machines regression
Artificial neural networks
Time series analysis

R E S E A R C H   A R T I C L E

A B S T R A C T

Artificial intelligence machine learning has become very popular in recent years. It offers the ability to combine machine learning theory with many analyses such as classification, prediction models, natural language processing. Carbon dioxide emission is defined as the release of carbon, often caused by human nature, into the atmosphere. In the 19th century, the industrial revolution took place and the use of coal-powered industrial vehicles increased the amount of carbon released into the atmosphere. These gases released into the atmosphere have brought climate problems in proportion to the increase in temperature. Because of climate problems, the sweet water source of the earth's ice pack continues to melt and the sea level rises. Therefore, the amount of carbon dioxide emission (metric tons per person) Artificial Neural Networks (ANN), Support Vector Machines Regression (SVMR), estimated by Box-Jenkins technique based on time series analysis and estimated estimates compared to MSE (mean square error) between 1990-2018. The comparison found that the Artificial Neural Networks have better predictive results on the SVMR and Box-Jenkins technique on the performance benchmark.

Turkish Journal of Forecasting by Giresun University, Forecast Research Laboratory is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

1. Introduction

Problems such as global warming and environmental deterioration are increasing day by day since the beginning of 1990. Global warming and climate degradation are the most important cause of increased emissions of carbon dioxide in the air. Economic resources and modelling of carbon dioxide emissions are among the issues researchers have been focusing on for years. When the literature is examined;

Ramanathan (2006) conducted Data Envelope Analysis (DEA) with links between CO_2 emissions, GDP growth, and energy consumption from 1980-2001. Carbon dioxide emissions are limited to levels released in 1990 and are determined by the estimation model under the highest value assumption for 1980 of the efficiency index in 2025. The model stipulated for 2025 states that the non-fossil energy consumption required to meet the GDP level will be much smaller than the values recorded in 1990 (44.59). However, the non-fossil energy consumption (118.8) in 2025

* Corresponding author.
E-mail addresses: fcemrek@ogu.edu.tr (Fatih Çemrek)

https://doi.org/10.34110/forecasting.1035912
2618-6594/© 2021 Turkish Journal of Forecasting. All rights reserved.

has shown that the efficiency index in 2025 is at the level recorded for 1990 and that the activity recorded in 1990 will increase [1].

Keskingöz et al. (2015) between 1960-2011, Turkey's foreign trade, energy consumption, and economic growth investigated the impact of CO_2 emissions. The approach to the ARDL Neural test is used as a method. As a result of the study, they found the presence of CO_2 emissions, foreign trade and growth, and long-term relationships [2].

Chang (2015) have created a unified connection analysis and multi-purpose programming structure to determine the production structure optimized for the industries producing CO_2 in China and the emission reduction target. In order to reduce CO_2 emissions from 5707.16 to 5452.12 million tons as a result of their work, China has been suggested to change its industrial structure by focusing on the industrial groups defined by the connectivity features. They said this would lead to an 82.59 billion Yuan GDP drop in China [3].

Pabuçcu et al. (2016) have worked on a model that predicts greenhouse gas oscillation values in the future with EU-28 countries. Between 1990-2030, they used population, GDP, energy production and consumption, energy usage for transport, and greenhouse gas release quantities. The method uses the Artificial Neural Networks Model (ANN). The forecast model found 740,33 million tons (MT), 1039,32 MT, and 1244,13 MT with a row of CO_2 equivalent oscillation quantities for 2020-2025-2030 [4].

İnce et al. (2016) discussed the supplier selection issue in their work. In their work, they used support vector regression (SVR) and twin support vector regression (TSVR) methods to predict the supplier credit index. Predictive precision has been compared between SVR and TSVR techniques in the determination of suppliers, and actual examples have shown that the TSVR technique is better for the SVR technique [5].

Chen et al. (2018) estimated carbon dioxide (CO_2) flux emissions with published data on the reserve of freshwater sources. They used the Artificial Neural Network (ANN), the back-spread Neural Network (BPNN), and the Generalized Regression Neural Network (GRNN) in their work. The results have been compared with regression models and ANN and have shown that ANN results more accurately than regression models and that GRNN provides the best performance. With the help of this GRNN, the total CO_2 emitted by global reservoirs has been estimated and possible CO_2 flow emissions have been evaluated from a planned reservoir showing the potential application of GRNN [6].

Chiu (2020) developed a new range of variable gray predictions models (MGPM) for CO_2 emissions. Compared to other models, MGPM's have stopped at several distinguishing features of the proposed model. First, both feature selection, and second, where changes are considered, a neural network-based residue model is used. They have shown that the proposed MGPM performs better than the experimental results [7].

Keskin (2020) during the period of 1960–2012, studied carbon dioxide emissions in the developed G-20 community, the relationship between national income per capita, co-integration variance parsing, and linear regression models. It has been concluded that CO_2 emissions have travelled on a yacht over the years [8].

Açıkkar et al. (2020) the upper thermal value of Turkish lignite charcoal washed in its work (GCV) was estimated using machine learning methods and dry basic short analysis results of coal samples. In addition, they implemented three different Artificial Neural Networks (ANN) in their work, including Support Vector Regression (SVR) and Multilayer Sensor (MLP), General Regression Neural Network (GRNN), and radiant-based function Neural Network (RBFN), and GCV estimation models. Predictive models are compared to performance measurements. For GCV estimation, all models indicate that the SVR-based model is better than ANN-based models and THE RBFN-based model performs better than MLP and GRNN-based models [9].

2. Data and Method

2.1. Data

The data set used in the study consists of CO_2 emission (metric per capita, ton), GDP (current USD per capita), population, Energy consumption (Metric tons per capita), agricultural land (hectares per capita) in Turkey between 1990-2018. Data from the World Bank and TUIK.

- Slave variable (target): CO_2 emission
- Arguments: GDP, population, Energy use, Agriculture

In the method of artificial neural networks and support vector machines, data has been normalized in the range 0-1 for the consistency of data within one another. The standardization of data is as in Equality 1.

$$X_n = X - X_{min} / X_{max} - X_{min} \tag{1}$$

Data logout has been taken for the analysis of time series. Naming data is shown in Table 1.

Table1. Name Data

Em	CO_2
Et	Energy consumption
ta	Agricultural land
n	Population
GDP	Gross Domestic Product

2.2. Artificial Neural Networks

Artificial neural networks are a computing technology that is developed by mimicking the way the human brain processes information. It consists of modelling the electrical connection between artificial neural networks and biological neurons and those cells in the computer environment. The nerve cells in the cells are made up of cores and axons. Biologically, the core provides a flow of information along the axons. The sensor data from the output terminal dendrite end is weighed in the core and connected to another neural cell along the axon. This way, neural communication is occurring. The mathematical model of a neural cell in the human is shown in Figure 1 and Figure 2.

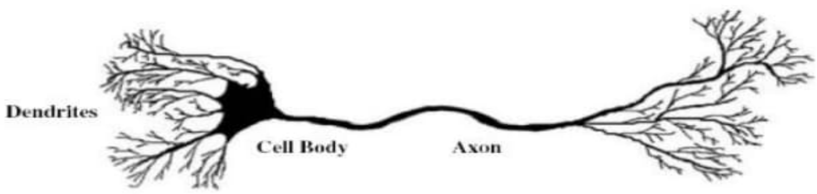

Figure 1. Human neural network [10]

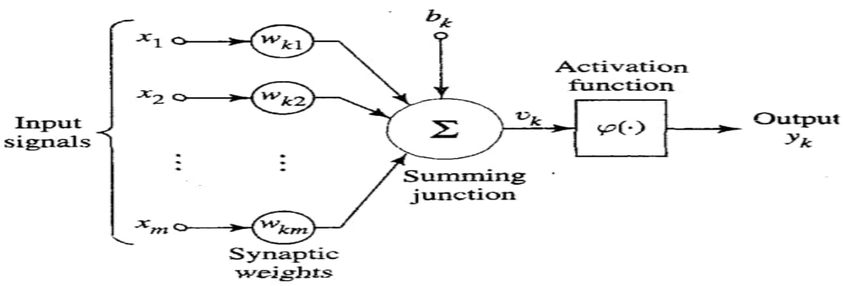

Figure 2. Mathematical model of the border cell in man [10]

There are road-length weights called dendrites, and we (x_{m}, $m = 1,2 \ldots \ldots n$) have input values that may have come from another neuron entering these dendrites. Input values and counterweights (w_{k}, $k = 1,2 \ldots \ldots n$) are transmitted to the Neural cell after multiplying ($w_k x_m$). Weight from all dendrites00 and input multiplication are collected. It is then b_k transmitted to the output via the activation function after it has been collected with a bias. This output may be the last output cell, but it may be at the entrance of another cell. Mathematically, weights are multiplied by inputs and collected with lice. This results in a mathematical model at the basic level. The actual action is to calculate the *w* (weight parameter) and *b* (bias value) parameters that the model will give the optimal value. Learning on Artificial Neural Networks is carried out using the controlled learning method. Practice by providing input-output information during learning.

Classical statistical methods are not suitable for data that is missing and/or excessive deviation due to the risk of incorrect results. The artificial neural network approach is independent of data; they can assess missing, partially incorrect, or excessive deviated data and learn, generalize complex relationships, and thus create a solution with a specific error that is acceptable to questions they have never encountered before. Due to these characteristics, artificial neural networks are used as an effective technique in estimating [11].

The statistical provisions of the terminology used in ANN are given in Table 2 below. It is important to be able to see the statistical basis.

Table 2. Statistical provisions of the termination line

TSA terminological expressions	Statistical terminological statements
Artificial Neural Network	Algorithm (model)
Weight (w)	Coefficient
Input (x_m)	Independent variable
Output (y_k)	Prediction value
Target (y)	Dependent variable
Error	Now (e)
Error line	Confidence range

Artificial neural networks are used in areas such as primary detection, classification, estimation, control, data filtering, etc., and prediction modelling has been used in many studies in recent years. It is not stated in the literature that it is very effective in complex problems**.**

2.3. Support Vector Machines (SVM)

Support vector machines are a supervised learning algorithm based on statistical learning theory based on the foundations of Vladimir Vapnik and Alexey Chervonenkis in 1963. In 1995, Bernhard Boser and Isabelle Guyon developed Vladimir Vapnik [12]. Support Vector machines are mainly used to best separate data from each other of the two classes. The hyperplane is determined for this.

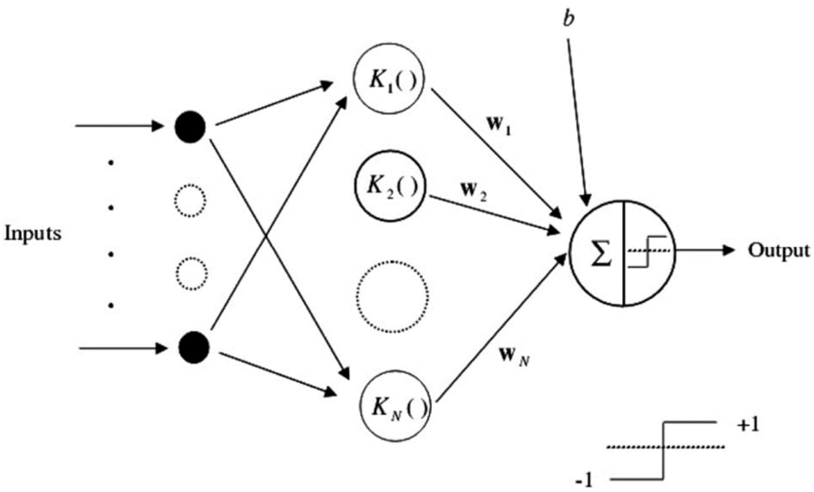

Figure 3. Support vector machines network structure [13]

2.4. Support Vector Machines Regression (SVMR)

In 1997, Vapnik and his friends to include regression practices expanded the SVMR algorithm [14]. The regression uses an “approach error” instead of the margin between an optimal splitter hyperplane and support vectors. As with the classification algorithm, the regression problem is tried to find optimal generalization at the top of the border. [15].

The structural risk in the Support Vector Regression includes two parts, the empirical risk of the function and complexity, and the equation is written as in 2.

$$R_{reg} = CR_{amp}[f] + \frac{\|w\|^2}{2} = C\sum_{i=1}^{l}|y_i - f(x_i)|_{\in} + \frac{\|w\|^2}{2} \qquad , x_i \in R^n , y_i \in R , w \in R^n \tag{2}$$

The learning situation here is the finding of the f (x, w0) formula, which minimizes the risk function. The risk for regression is a combination of noise variance and accuracy approach. The least risky prediction is equivalent to finding the most accurate unknown g(x) function.

The steps to follow in the SVMR training are as follows;

Step 1: Determine the Kernel function which forms the regression function in the regression problem
Step 2: Parameter selection of the selected Kernel function
Step 3: Choice of the punishment parameter C
Step 4: The solution to the second-degree problem.

2.5. Box-Jenkins (ARIMA) Models

Statistician George Box and Gwilym Jenkins developed the ARIMA models. These models apply the mean or mobile models with authoritarian movement or authority for values that correspond to the historical time data of the time series. The box-Jenkins method refers to the fact that each time series has a function of historical values and can be explained with them. This technique shows a successful performance in the short-term estimate. The fact that the series in which the technique is applied is a series of intermittent and stationary, with equal time intervals and observation values is one of the key assumptions of this method [16].

Box-Jenkins models are examined in three groups and are:

- Linear stationary scholastic models,
- Non-stationary linear stockpiled models,
- Seasonal models.

The stages of the Box-Jenkins model; setting the appropriate model consists of the model's prediction, inspection, and prediction.

3. Application and Findings

3.1. Artificial Neural Networks Results

In the creation ANN model, a forward-feed reverse-propagated neural network algorithm is used in the literature frequently. In practice, we use this algorithm because it has low predictive performance, ease of use, and convergence time in linear and nonlinear models. When making multi-term estimates with ANN; Two different approaches can be used, a single-term, successive linear system and a direct technique where multiple semesters are predicted at the same time [17]. The design of the model is done with the Matlab software program toolbox, and all steps, such as the training, testing of the model, and prediction production of the model, have been done with this software. Data set: the training is divided into three sub-sets, verification, and testing. In assigning data to sub-groups at the specified rates, random methods were applied from random and sequential techniques. As a learning algorithm, Scott E. from Carnegie Mellon University Fahlman developed and used for general purpose Quick Propagation has been selected. The Fast Smear algorithm is an intuitive learning algorithm used to train Newton technique-based and multi-layered neural networks [18]. In creating the optimal network architecture, the continuous and nonlinear logistics (sigmoid) function is preferred, which is the common use of input and output layers within the activation functions. The input layer of the network has four inputs (input) of four independent variables available to the network, and one output (output) of the dependent variable in the output layer. It has been decided to use ten hidden layers because of attempts to determine the number of hidden layers and the number of neurons to be included in the hidden layer.

Figure 4 shows four input variables (arguments), ten hidden layers, one output result, and the Neural network model.

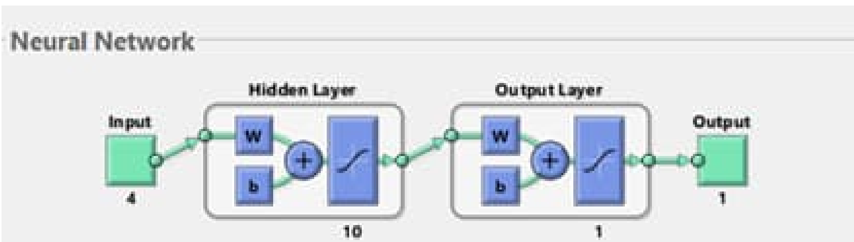

Figure 4. Created artificial neural network model

The step performance criteria of the artificial neural network was given in Table 3.

Table 3. Artificial neural network results

Iteration	R	MSE
0	0.99885	0.00017598
78	0.99746	0.00036085
105	0.99771	0.00047657
155	0.99783	0.000069329
967	0.99914	0.000023727***

Table 3 the best prediction model, when reviewed according to the MSE (Mean Squared Error) performance criteria, is 967. It appears in the iteration. The correlation diagram of the model that performs the appropriate value selected is given in Figure 5.

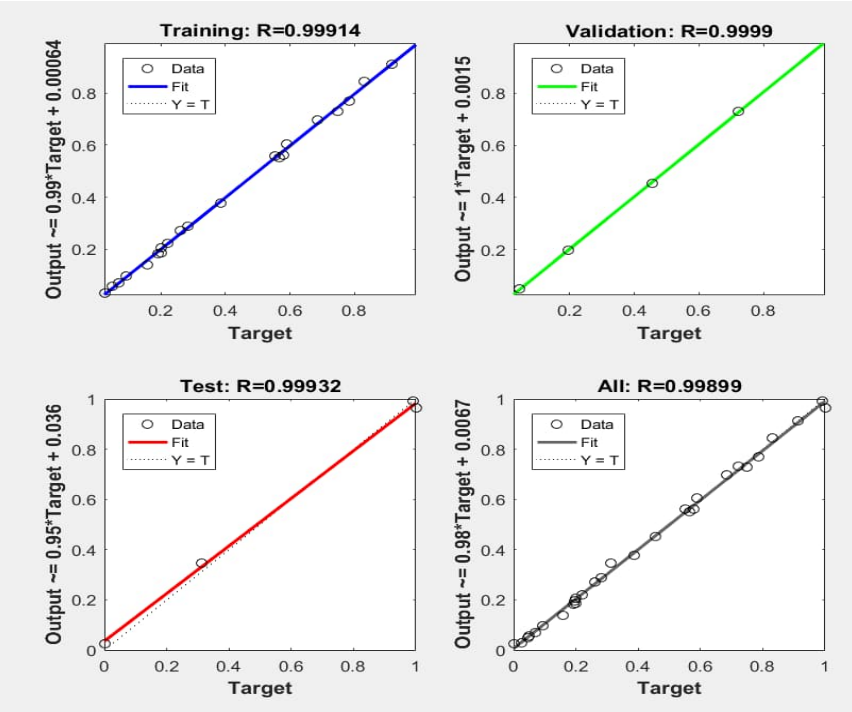

Figure 5. Correlation graphics of selected artificial neural network model

From the terms mentioned in Figure 5, the target variable, output estimates, test training data, and validation statement mean validity. In Figure 5, the correlation of the training data is 0.99 and the test data is 0.99.

3.2. Support Vector Machines Regression (SVMR) Results

The C = [0.1, 0.5, 1, 0.6] matrix is used for the penalty parameter. Because of matrix optimization, the penalty parameter is selected as 0.5. The estimated regression equation because of six cross-verifications is as in Equation 3.

$$\hat{Y} = 0.25417093+0.25275792*Et + 0.26660736*GDP +0.24534774*n -0.24789029*ta \quad (3)$$

The performance criteria for the predicted regression equation MSE = 0.03671903 has been found.

3.3. Box-Jenkins Model Results

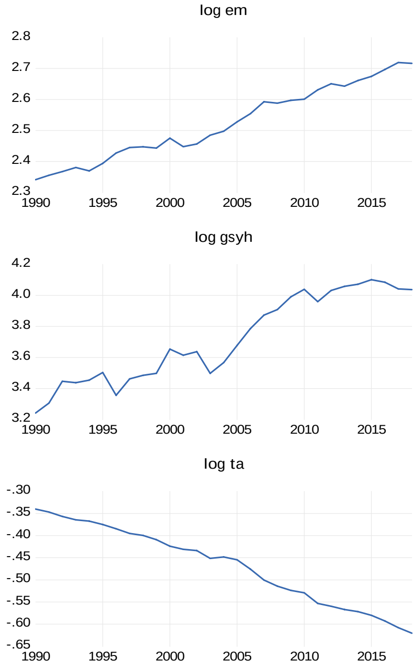

Figure 6. Time series graph of logarithmic data

CO_2 emission and other data are logarithmic and stagnated with a level difference. As the appropriate model, ARIMA (0, 1, 0) is specified, and the performance values are given in Table 4.

Table 4. Model result for ARIMA

ARIMA Model	R^2	MSE
(0,1,0)	0.977	0.0144
(1,0,0)	0.957	0.0201
(1,0,1)	0.935	0.0182

ACF and PACF graphs of the selected model are given in Figure 7.

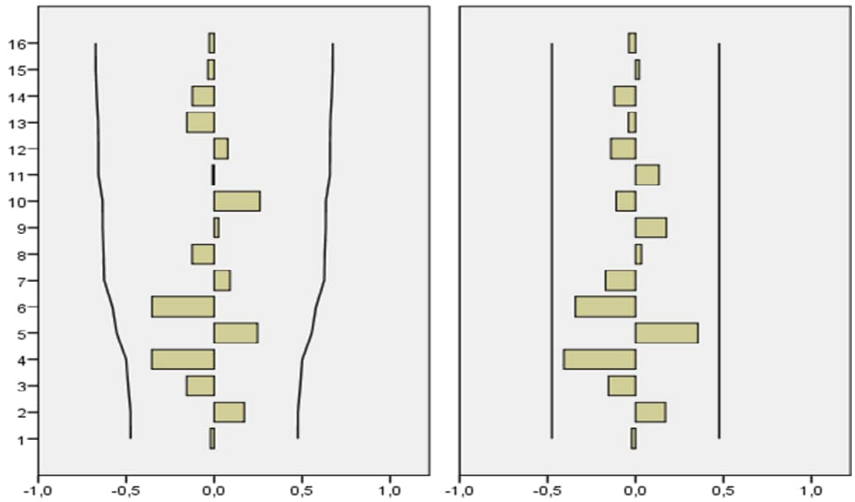

Figure 7. ARIMA (0,1,0) ACF and PACF

3.4. Benchmarks and Estimation Accuracy

MSE (mean square error) is used as a common benchmark. When we look at the individual findings, it is taken into account in the R's which provide the described part of the argument. Performance criteria are given in R Equation 4 and MSE Equation 5.

$$\text{MSE}=\frac{1}{n}\sum_{t=1}^{n} e_t^2 \quad (4)$$

$$\text{R}=\frac{\sum_{i=1}^{n}(\hat{Y}_{\text{i}}-\bar{Y})^2}{\sum_{i=1}^{n}(Y_{\text{i}}-\bar{Y})^2} \quad (5)$$

Table 5. Model and performance criteria

Model	MSE
SVMR	0.03671903
ANN	**0.000023727****
BOX-JENKINS	0.0144

Table 5 the most successful prediction performance is seen as artificial neural networks (MSE=0.000023727) from machine learning methods.

4. Results

In this study, Support Vector Regression, Artificial Neural Networks, and Box-Jenkins method from machine learning techniques were applied to model CO_2 emissions. Methods were compared in terms of MSE criteria. Compared to the performance of the projections obtained from every three methods, ANN has a better prediction performance, with better results than SVMR and Box-Jenkins. The best predictive model of algorithms from methods is the model of artificial neural networks. In the support vector regression, data is taken into account based on a specific parameter and this parameter ensures that the data to be taken into account is at least epsilon from the regression line to be created, and therefore the support vector regression algorithm does not provide optimal results in the probability processes. This result is Jain et al. (2001), Bougadis et al. (2005), Msiza et al. (2008), and Adamowski (2008) are supported by their work [19-22].

We have seen better predictions than classical time series analysis can be done with machine learning algorithms. Our work, prepared to compare new techniques and practices for modelling CO_2 emissions, a cause of global warming, the biggest problem in today's world, has been written to contribute to the literature. It is recommended to increase coverage for the most appropriate model selection by applying different machine learning methods for CO_2 emissions modelling.

References

[1] Ramanathan, R. (2006), "A multi-factor efficiency perspective to the relationships among world GDP, energy consumption and carbon dioxide emissions", Technological Forecasting and Social Change, 483-494.

[2] Keskingöz, Hayrettin, Karamelikli, Hüseyin, (2015), "Dış Ticaret-Enerji Tüketimi ve Ekonomik Büyümenin CO2 Emisyonu Üzerine Etkisi", Kastamonu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi 9(3), 7-17.

[3] Chang, N. (2015), "Changing industrial structure to reduce carbon dioxide emissions: a Chinese application". Journal of Cleaner Production, 40-48.

[4] Pabuçcu,Hakan., Bayramoğlu, Turgut, (2016), "Yapay Sinir Ağları İle CO2 Emisyonu Tahmini: Türkiye Örneği". Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 18(3), 762-778.

[5] İnce, H., ve İmamoğlu, S. (2016). Destek vektör regresyon ve ikiz destek vektör regresyon yöntemi ile tedarikçi seçimi. Doğuş Üniversitesi Dergisi, 17(2), 241-253.

[6] Chen Z, Ye X, Huang P. (2018), "Estimating Carbon Dioxide (CO2) Emissions from Reservoirs Using Artificial Neural Networks". Water. 2018, 10(1):26.

[7] Chiu, Y., Jiang, P., etc. (2020), "A Multivariate Grey PredictionModel Using Neural Networks with Application to Carbon Dioxide Emissions Forecasting". Hindawi.

[8] Keskin, A. (2020), "G20 Ülkelerinin CO2 Emisyonları ve Kişi başına Düşen Milli Gelirin 1960-2012 Dönemi Arasındaki Ampirik Analizi", Medeniyet Araştırmaları Dergisi, 5(1), 25-38.

[9] Açıkkar M., Sivrikaya O., (2020), “Yıkanmış Türk Linyit Kömürlerinin Üst Isıl Değerinin Destek Vektör Regresyonu ile Tahmini”, Avrupa Bilim ve Teknoloji Dergisi. 2020, (18), 16-24.

[10] Haykin, S., (1999), “Neural Networks: A Comprehensive Foundation”. 2nd Edition, Prentice-Hall, Englewood Cliffs, NJ.

[11] Özalp A., Anagün S., (2003). “Yapay Sinir Ağı Performansına Etki Eden Faktörlerin Analizinde Taguchi Yöntemi: Hisse Senedi Fiyat Tahmini Uygulaması”, İstatistik Araştırma Dergisi, Yıl: 2, Sayı: 1.

[12] Akpınar M., (2017), “İstatistik Ve Yapay Zeka Teknikleri İle Enerji Tüketiminin Tahmini: Sakarya Doğal Gaz Tüketiminin Uygulaması”, Doktora Tezi,Sakarya Üniversitesi Fen Bilimleri Enstitüsü.

[13] Uçak, K., (2012) “Destek Vektör Regresyonu ile Pid Kontrolör Tasarımı”. İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 89s.

[14] Bilişik, M. T. (2011), “Destek Vektör Makinesi, Çoklu Regresyon ve Doğrusal Olmayan Programlama ile Perakendecilik Sektöründe Gelir Yönetimi İçin Dinamik Fiyatlandırma”. İstanbul: XI Üretim Araştırmaları Sempozyumu.

[15] Walgampaya C., Kantardzic M. “Selection of Distributed Sensors for Multiple Time Series Prediction”, The 2006 IEEE International Joint Conference on Neural Network Proceedings.2006.

[16] Özmen, A. (1986), “Zaman serisi analizinde Box- Jenkins yöntemi ve banka mevduat tahmininde uygulama denemesi”.Kitap. Anadolu Üniversitesi Yayınları 207. Fen Edebiyat Fakültesi Yayınları No: 9.

[17] Karahan, M. (2011), “İstatistiksel Tahmin Yöntemleri: Yapay Sinir Ağları Metodu ile Ürün Talep Tahmini Uygulaması”. Yayınlanmamış Doktora Tezi. Konya: Selçuk Ü. Sosyal Bilimler Enstitüsü.

[18] Tozkan, S., (2004) “Yapay Sinir Ağları”, Fırat Üniversitesi Mühendislik Fakültesi Elektrik Elektronik Mühendisliği Bölümü Bitirme Ödevi.

[19] Adamowski, J. F. (2008), “Peak Daily Water Demand Forecast Modeling Using Artificial Neural Networks”, Journal of Water Resources Planning and Management, 134(2), 119-128.

[20] Bougadis, J., Adamowski, K., Diduch, R. (2005), “Short-Term Municipal Water Demand Forecasting”, Hydrological Processes. 19, 137–148.

[21] Msiza, I. S., Nelwamondo, F. V., & Marwala, T. (2007), “Water Demand Forecasting Using Multi-layer Perceptron and Radial Basis Functions”. Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, USA. August, 12-17.

[22] Jain, A., Varshney, A. K., Joshi, U.C., (2001), “Short-Term Water Demand Forecast Modelling at IIT Kanpur Using Artificial Neural Networks. Water Resources Management. 15, 299-321.