Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2024) 32: 51 – 67

doi:10.55730/1300-0632.4055

Artificial intelligence-based evaluation of the factors affecting the sales of an iron and steel company

Mehmet PEKKAYA[1], **Zafer UYSAL**[1], **Aytaç ALTAN**[2*], **Seçkin KARASU**[2]

[1]Department of Business Administration, Zonguldak Bülent Ecevit University, Zonguldak, Turkiye

[2]Department of Electrical Electronics Engineering, Zonguldak Bülent Ecevit University, Zonguldak, Turkiye

Received: 07.08.2023 • **Accepted/Published Online:** 08.12.2023 • **Final Version:** 07.02.2024

Abstract: It is important to predict the sales of an iron and steel company and to identify the variables that influence these sales for future planning. The aim in this study was to identify and model the key factors that influence the sales volume of an iron and steel company using artificial neural networks (ANNs). We attempted to obtain an integrated result from the performance/sales levels of 5 models, to use the ANN approach with hybrid algorithms, and also to present an exemplary application in the base metals industry, where there is a limited number of studies. This study contributes to the literature as the first application of artificial intelligence methods in the iron and steel industry. The ANN models incorporated 6 macroeconomic variables and price-to-sales data and their results were evaluated. An ordinary least squares regression model was also used to facilitate the comparison of results, while gray relational analysis (GRA) was used to draw a comprehensive conclusion based on the ANN results. The results showed that the variables USD/TL exchange rate, product prices, and interest rates, in descending order, had the highest degree of influence in determining the sales of the iron and steel company. Furthermore, these variables are crucial for forecasting future sales and strategic planning. The study showed that the ANN outperformed classical regression models in terms of prediction accuracy. In the model applications conducted for 5 different product groups, it was observed that 3 models (models 2, 3, and 4), including model 4, which sold a higher volume of products than the total of the other products, had an overall performance above 80%. In addition, GRA was found to be a valuable tool for synthesizing insights from different ANN models based on their respective performance levels.

Key words: Iron/steel sale determiners, artificial neural network, gray relational analysis, deriving integrated inferences

1. Introduction

Forecasting an iron and steel company's future sales is a critical decision that affects many aspects of the business, from plant operations to production planning. Projected sales directly correlate with expected operating income, as income is largely derived from sales. The iron and steel industry serves as a primary source of raw materials for manufacturing, supplying various product groups that act as inputs for various industries. Therefore, the health of the iron and steel sector has a significant impact on related subindustries and contributes to overall national development. A slowdown or stagnation in this sector can serve as an indicator due to its interconnectedness with other industries. Given the large scale and high tonnage production that characterizes the industry, the

*Correspondence: aytacaltan@beun.edu.tr

accuracy of sales forecasts has significant financial implications. For example, a large iron and steel company may experience sales fluctuations of 200,000 to 300,000 tonnes per year for a single product with an average price of 700 USD per tonne. A 20% error in demand estimation for just one product translates into a variation in revenue of approximately 35 million USD [1].

Türkiye has 34 companies engaged in crude steel production, with three integrated plants. In 2021, Türkiye was the seventh largest steel producer in the world, contributing 40.36 million tonnes to the global production of 1.951 billion tonnes (an increase of 12.7% on the previous year). It is also the second largest steel producer in Europe after Russia. Steel exports amounted to 22.4 billion dollars in 2021, accounting for 9.9% of Türkiye's total exports [1]. Given the size of the iron and steel market and its links with other industries, forecasting future sales is becoming increasingly important.

Market fluctuations, influenced by economic and political conditions, can affect sales in certain periods. Economic downturns, recessions, or abrupt changes in macroeconomic indicators can affect different industries. Understanding the factors that influence product demand during changes in the external economic climate and assessing their impact play a key role in sales forecasting. While numerous academic studies have examined sales/demand forecasting in various industries such as energy, oil, tourism, retail, and housing, research on this important economic issue in the metals and steel industries remains scarce, which underscores the importance of our study. The motivation for the present study is the search for a method/model that can better predict the sales of companies in the iron and steel sector, which provides basic inputs to almost all industries, and the need to know the priorities of variables that must be taken into account in forecasting the sales of companies in the sector in order to ensure better planning, and the fact that there are very limited studies in the field that include these purposes. The aim in our study was to prioritize and assess the factors influencing the sales forecasts of major iron and steel companies using an effective methodology.

In the field of time series data, including financial, economic, energy, weather, and stock prices, traditional time series methods as well as artificial neural networks (ANNs) and machine learning techniques are widely used to accurately predict future values [2–6]. ANNs provide nonlinear models that excel at capturing nonlinear relationships between input and output data, making them superior for prediction [7]. ANNs use the backpropagation algorithm during training, which aims to minimize the mean squared error between expected and actual outputs across the training dataset. This supervised training technique allows designers to produce the desired result. Although many studies on sales forecasting can be found in the literature, these studies are mainly conducted in the energy, tourism, retail, housing, and oil-related sectors. The number of studies in the metal sector is quite limited, and compared to the studies observed, almost no study has been found in which as many important variables as in our study have been evaluated with hybrid ANN approaches and the integrated results of the variables affecting sales from models. These statements set out the contributions of our work to the literature.

The primary objective of the present study was to identify and model the significant factors influencing the sales of an iron and steel company using an ANN. Subsequently, we aim to present these results to decision makers and researchers. We first selected the key factors based on the relevant academic literature. In addition, we selected the 5 best-selling products from 5 product groups produced by a large Turkish iron and steel company and collected monthly sales data for the last 10 years. After collecting the variable data, we constructed ANN models using the Levenberg–Marquardt training algorithm and evaluated the results. To compare the

[1]TSEA (2022). Turkish Steel Exporters' Association [online] Website http://www.cib.org.tr/tr/istatistikler.html [accessed 03 August 2023]

performance of the ANN models, we built an ordinary least squares regression model to determine the variables that significantly affect sales. We then statistically compared 1 year of data (the test period) with the previous 9 years of data in the ANN model.

In the second section, we assess the iron and steel industry in Türkiye and worldwide, focusing on production and market size. In the third section, the factors influencing the demand for iron and steel are identified based on a literature review. The fourth section contains details on the aforementioned data periods, variables, and analysis methods, as well as the results of the application. Finally, in the last section, the results are evaluated.

2. Iron and steel industry in Türkiye and the rest of the world

The iron and steel industry is regarded as a driving force in the industrialization of nations and the advancement of their economies. Its significance lies in its role as a supplier of raw materials to various industrial sectors. An examination of the relationship between a country's development and the iron and steel industry reveals that advances and innovations in this sector contribute significantly to the growth of iron- and steel-related industries, thereby having a multiplier effect on a nation's development. The iron and steel industry supports the production of transportation vehicles, including automobiles, ships, containers, aircraft, railways, and wagons. It also plays a crucial role in the production of packaging materials, machinery, agricultural tools, construction materials, heating equipment, and household and office goods [2].

Crude steel production in Türkiye is carried out in both arc furnace facilities and integrated facilities. Integrated plants utilize iron ore as their primary raw material, while arc furnace plants rely on iron and steel scrap. Approximately one-fourth of Türkiye's total crude steel production is derived from integrated facilities, with the remaining three-fourths originating from arc furnace facilities. The prevalence of arc furnace facilities in the industry is attributed to the substantial investments and financing required for integrated facilities, as well as the limited availability of iron ore deposits in Türkiye. The iron and steel industry provides employment to around 88,000 individuals within various companies [2].

In 2021, Türkiye was the seventh largest steel producer (Table 1) in the world and the second largest in Europe. Moreover, iron and steel exports accounted for 9.9% of Türkiye's total exports in the same year. Given the important role of the iron and steel industry in the Turkish economy and its impact on various other sectors, the aim in our study was to model the factors influencing future sales volumes in this industry. Accurately estimating sales for future periods will further underscore the relevance and impact of our research.

3. Literature review on the factors determining the sales of and demand for iron and steel

Academic studies on the factors influencing the demand forecast, sales forecast, and consumption forecast in the iron and steel industry were searched. It is stated that the economic development of countries is related to steel consumption and it can be regarded as an indicator of the development of the country's industry [8]. In [9] it is suggested that British steel demand is sensitive to the intensity of steel use and to changes in the level of macroeconomic activity. It supports the view that economic activity affects steel demand and uses the variables of national income, exchange rate, steel price, and state of the manufacturing industry, which are among the economic factors affecting steel consumption [10]. Apart from these, in the Japanese steel industry [11], in the Polish steel industry [12], in the Chinese steel industry [13], in the British steel industry [14], and in the Korean

[2] İhracat Genel Müdürlüğü Maden, Metal ve Orman Ürünleri Daire Başkanlığı (2017). Demir-Çelik, Demir-Çelikten Eşya Sektör Raporu [online]. Website https://www.orhangazitso.org.tr/webFiles/1488897357.pdf [accessed 03 August 2023]

steel industry [8], the effects of economic atmosphere, national income, and the intensity of use of steel in the industry on steel consumption were examined in his studies of the UK steel industry.

Table 1. World crude steel production 2021 (thousand tons)[3].

Order	Country	2021	Rate (%)
1	China	1,032,790	52.9
2	India	118,244	6.1
3	Japan	96,334	4.9
4	USA	85,791	4.4
5	Russia	75,585	3.9
6	South Korea	70,418	3.6
7	Turkey	40,360	2.1
8	Germany	40,066	2.1

Order	Country	2021	Rate (%)
9	Brazil	36,174	1.9
10	Iran	28,460	1.5
11	Italy	24,426	1.3
12	Taiwan	23,233	1.2
13	Vietnam	23,019	1.2
14	Ukraine	21,366	1.1
15	Mexico	18,454	0.9
	World	1,951,924	100.0

Table 2. Factors determining sales/demand.

Researcher	Subject	Variables	Method
[10]	UK steel demand	National income, interest rate, steel price, steel-related industry demand	Vectorial regression
[11]	Future trends in Japanese steel consumption	National income, population, steel-related, industry demand	Usage intensity model
[15]	Printed circuit boards sales forecast	National income, consumer price index, manufacturing product index	Genetic algorithm, feedback ANN
[16]	Electricity demand in Athens/London	National income, demographic structure, air temperature	Trend analysis, seasonal analysis
[17]	White goods supply chain demand forecasting	Product price, product quality, promotions	Fuzzy logic, ANN
[13]	Consumption forecast in the Chinese steel industry	National income, per capita income, intensity of use	Bayesian vectorial autoregression
[8]	Steel consumption and long-short-term economic development in Korea	National income, per capita income	Vectorial autoregression
[18]	Packaging sales forecast	Manufacturing consumer index, competitive index, historical sales data	Delphi and ANN
[19]	Electricity consumption forecast	Population growth rate	Vector regression supported ANN
[20]	Demand forecasting in the iron/steel industry	National income, national income growth rate, inflation, steel production data	Fuzzy logic, ANN
[21]	Factors affecting energy demand in developing countries	National income, price, economic structure, CO2 emissions	Dynamic panel analysis
[12]	Demand forecast in the iron-steel industry based on the business climate index	National income	ARIMA, SARIMA
[14]	Approach to determining forecast parameters: An application to steel intensity of use in the UK	Intensity of use	Grey Verhulst model
[22]	Automotive demand forecasting	Historical sales data	ANN
[23]	Sales forecast of a company	National income, exchange rate, inflation, historical sales data	ANN with fuzzy logic revision
[24]	Stainless Steel sales forecast	Raw material prices, USD/TRY, PPI, industrial production index	Data mining, model tree method

Given the limited number of studies specifically looking at factors influencing product sales and demand, we broadened our perspective by including findings from different industries, as summarized in Table 2. The table shows that certain economic factors are commonly used in academic research on sales. These factors

[3]WSA (2022). World Steel Association [online] Website http://www.worldsteel.org [accessed 03 August 2023]

include national income, the manufacturing index, inflation, exchange rates, product prices, and historical sales/production data. In this research endeavor, we consider these variables as our primary focus.

Traditional forecasting techniques such as trend analysis and moving weighted averages, which are commonly used in sales forecasting, are generally recognized for their cost-effectiveness and ease of use. However, they may fail to adequately take into account the passage of time, seasonal influences, and economic fluctuations in their forecasts. Economic markets can fluctuate based on a country's broader economic and political landscape, which can subsequently affect sales. Economic downturns, recessions, or shifts in macroeconomic indicators can have varying effects on product demand within specific industries. Identifying these factors that shape product demand in response to changes in the external economic climate and assessing their impact play a key role in forecasting future sales.

The literature on sales forecasting encompasses a wide array of scientific studies spanning different industries, including energy, oil, tourism, retail, and real estate. The fact that few studies have been conducted on the Turkish metal industry or the iron and steel industry increases the importance of the present study. As a result, research on the economic factors influencing the turnover of companies in the iron and steel industry, which is the focus herein, is of paramount importance.

4. Materials and methods

The aim of the study was to determine and model the significant factors with their priorities that influence the sales volume of an iron and steel company by means of an ANN. In this section, the processing map of the ANN model is shown in Figure 1. The following subsections explain the variables and data used in the application and the ANN model.

The flowchart illustrating the approach for modeling the significant factors influencing the sales of an iron and steel company using an ANN is presented in Figure 1. The first step is to prepare a product database to make it suitable for analysis. As a preprocessing step, separate datasets are created for 5 different products. Data imputation is performed using the k-nearest neighbors (KNN) technique. A min–max scaling approach is used to normalize the features and bring them into a uniform range. Subsequently, an ANN model is employed to predict the sales figures for each product. Finally, the model's performance is evaluated using test data. A detailed explanation of the ANN model can be found in section 4.2.

4.1. Data

As the analysis of iron and steel sales/demand is carried out for a Turkish company, Turkish macroeconomic variables are used in the modeling process. In line with the relevant literature, 6 economic variables that are assumed to affect iron and steel sales/demand are used, as shown in Table 2. In addition, the prices of the product and its lagged sales volumes are considered as potential factors influencing iron and steel sales/demand. These economic variables consist of the producer price index (PPI), gross domestic product (GDP), interest rate, US dollar to Turkish lira exchange rate (USD/TRY), consumer confidence index (CCI), and Purchasing Managers Index (PMI) for inflation (as shown in Table 3).

As the iron and steel industry has an impact on other industries, including manufacturing, some studies have often used the manufacturing index or the industrial index of the industries that affect the steel industry [10, 13, 15, 18, 24]. The PMI can be used to track the direction of manufacturing industries that use steel products as inputs. The PMI can be calculated using data from a survey of purchasing managers who are asked in which direction (increase/improve or decrease/worsen) parameters such as production, new orders, stocks, employment, supplier performance, and price trends are moving. As products are thought to affect the end user, the value of the manufacturing consumer index was also included in [18]. The consumer confidence index is also included in the model as it will show the tendency towards consumption of final products.

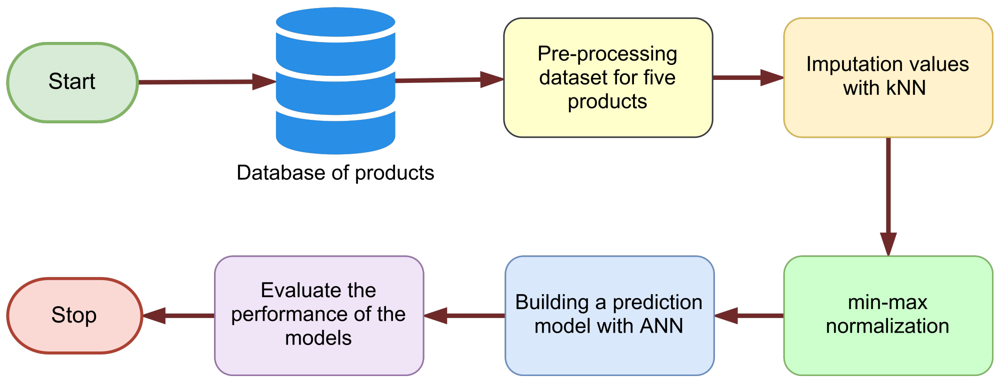

Figure 1. Flowchart of the modeling approach of the significant factors affecting the sales of an iron and steel company via an artificial neural network.

Table 3. Variables for models used in the study.

Variables	Used variable description	Start date	Data source
Q_i	Q_1, Q_2, Q_3, Q_4, Q_5 monthly average product sales (ton)	Jan. 2010	The company
P_i	P_1, P_2, P_3, P_4, P_5 monthly average product prices (USD)	Jan. 2010	The company
PPI	Producer price index (%) [inflation]	Jan. 2010	CBRT [4]
GDP	Gross domestic products [national income]	Jan. 2010	TSI [5]
$IntR$	CBRT interest rate	Jan. 2010	CBRT [4]
USD	USD/TL rate, Forex sale [exchange rate]	Jan. 2010	CBRT [4]
CCI	Consumer confidence index	Jan. 2012	TSI [5]
PMI	Purchasing managers index [manufacturing industry index]	Apr. 2015	ICI [6]

The variables of exchange rate [23], inflation rate [20, 23, 24], interest rate [10], product prices [10, 17, 21], and population data [11, 19] are economic factors affecting sales according to some studies. Accordingly, monthly producer price index as inflation data, monthly USD/TRY exchange rate, and market benchmark interest rates are used in the model. As the application is on manufacturing then the producer price index variable is used for the inflation variable in the analysis. Since the iron and steel trade is usually conducted in USD, then the USD/TL rate is used for the exchange rate. The PMI is regarded as equivalent to the manufacturing industry index.

It is also worth noting that in many of the studies listed in Table 2, national income, per capita income, and historical sales data have been commonly employed. Nevertheless, during data collection, it was observed that monthly sales and price data were available, while national income data were only accessible on a quarterly basis. To overcome this discrepancy, a linear interpolation method was applied to the national income data, allowing them to be incorporated into the model on a monthly basis. The monthly datasets for these variables covering the period from January 2010 to March 2021 were obtained, the sources of which are detailed in Table 3.

[4]CBRT (2021). Central Bank of the Republic of Türkiye [online]. Website https://www.tcmb.gov.tr [accessed 03 August 2023].
[5]TSI (2021). Turkish Statistical Institute [online]. Website https://www.tuik.gov.tr [accessed 03 August 2023].
[6]ICI (2021). Istanbul Chamber of Industry [online]. Website http://www.iso.org.tr/ [accessed 03 August 2023].

In the analysis, we scrutinized the sales data (Q_1, Q_2, Q_3, Q_4, and Q_5) of a Turkish iron and steel company for the period 2010 to 2021, covering about 11 years, across 5 selected products. It should be noted that in the Q_3-P_3 dataset, 5 observations from the first year were found to be missing, and no imputation method was applied to address this data gap. Furthermore, the CCI dataset is available from January 2012 onward, while the PMI dataset is available from April 2015 onward.

4.2. Artificial neural networks

ANNs are computer models that can synthesize and infer new information, use previously learned or classified information to make decisions, and mimic the organic neural structure of the human brain. The mathematical description of the learning process using the human brain as an example led to the development of ANNs. It resembles the organization of the brain's organic neural networks and their ability to learn, remember, and generalize. In biological systems, learning is achieved by modifying the synaptic connections between neurons. In other words, human learning begins at birth. The brain grows continuously during this process. Synaptic connections are modified or even created as a result of our experiences and daily life. This is how learning happens. ANN also fits into this. Training is the process by which learning by example occurs; in other words, realization occurs by processing input/output data, or more specifically, by the training algorithm continuously modifying the weights of the synapses using these data until convergence is achieved [25]. Calculating the bias value and w weight parameters, for which the model will provide the best score, is the fundamental step in ANNs. Neurons are connected to one another either in series or in parallel, and each neuron performs calculations in the same way. Five components make up an artificial neuron: inputs, weights, a summation function, an activation function, and outputs. Figure 2 shows these components for a simple single layer neural network. Inputs ($x_1, x_2, \ldots x_i$) are connected to neuron j in Figure 2, along with weights (w_{1j}, w_{2j}, ..., w_{ij}) for each connection. Each signal is amplified by the corresponding weights of the connection before being summed together by the neuron. The final output, o_j, is then obtained by passing this output through an activation function, which is often nonlinear [26].

5. Results and discussion

In this section, the results of the prioritization of the determining variables that influence the company's iron and steel sales using the ANN approach are presented. First, some descriptive statistics of the data are presented in Tables 4 and 5. One of the commonly known descriptive statistics, the coefficient of variation (CoV) shows the homogeneity of the series, which is preferred for the decomposition of the series from its mean and unit of measurement instead of the standard deviation [27]. CoV values can be calculated by dividing the standard deviation by the mean (CoV=100*S/Mean) of each series. According to the CoV values, PPI in particular, followed by *IntR* and *USD*, are more heterogeneous series than the others. In the following subsections, the model estimation is carried out using the ordinary least squares method for comparison with the ANN results. Then the ANN results are reported. Finally, the priorities of the variables obtained from ANN are evaluated comparatively using gray relational analysis (GRA). GRA is a multicriteria decision method (MCDM) and is used to draw an integrated conclusion from the results in different models of variable priority levels obtained from an ANN.

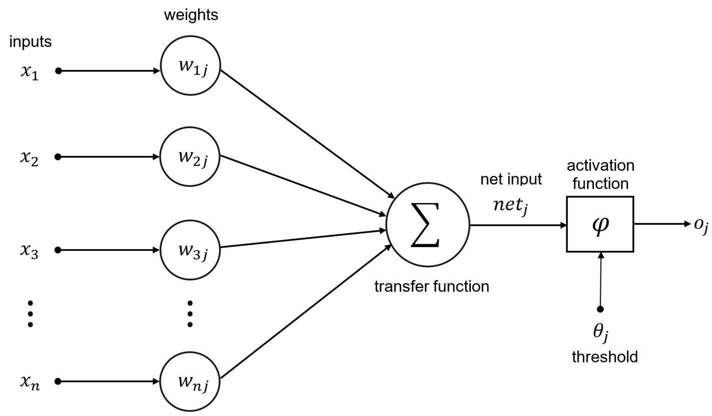

Figure 2. A single-layer ANN showing major components: inputs x_i, weights w_{1j}, summing function Σ, activation function φ, and output o_j.

Table 4. Descriptive statistics of the monthly average product sales and prices variables (S: standard deviation, CoV: coefficient of variation).

	Q_1	Q_2	Q_3	Q_4	Q_5	P_1	P_2	P_3	P_4	P_5
Min	24,132	7,088	6,690	61,260	2,882	263.0	295.0	310.0	129.0	334.0
Max	104,743	60,364	54,542	423,796	28,269	936.0	1,015.0	897.0	823.0	1,383.0
Mean	71,546	26,091	24,302	240,096	17,179	606.8	691.8	568.4	482.4	911.2
S	13,227	9,141	9,932	61,791	4,287	170.8	182.3	146.0	161.0	295.9
CoV	18.5	35.0	40.9	25.7	25.0	28.1	26.4	25.7	33.4	32.5

Table 5. Descriptive statistics of the common variables (S: standard deviation, CoV: coefficient of variation).

	PPI	$IntR$	USD	CCI	PMI	GDP
Min	1.7	4.5	1.4	76.9	33.4	152,268
Max	46.2	24.0	8.3	97.4	56.9	235,838
Mean	11.6	9.5	3.4	88.6	49.6	195,260
S	8.9	5.4	1.9	5.0	2.8	16,883
CoV	76.7	56.9	55.8	5.6	5.6	8.6

5.1. Model estimate results via ordinary least squares in order to compare ANN results

The ordinary least squares (OLS) method is employed as an alternative statistical approach to ANNs to identify the variables influencing iron and steel sales. OLS is a commonly used statistical method for estimating parameters, determining the impact of independent variables on the dependent variable [1, 28, 29], and predicting/forecasting unknown dependent values [30, 31]. This approach allows us to compare the results with those obtained using ANNs. Five separate regression models are constructed for the 5 products, using the same set of macroeconomic variables. The coefficients obtained from these models and their levels of significance are presented in Table 6. In order to address the issue of unit roots in the series, the regression is conducted using the logarithmic return series [32]. It should be noted, however, that the OLS assumptions are not met for these regression models. Specifically, it is noted that the normal distribution of the residual series is not satisfied and that there is an autocorrelation problem in all models, as indicated by the Breusch–Godfrey serial correlation LM test with a significance level of 0.05. Consequently, a heteroskedasticity and autocorrelation consistent (HAC) approach is applied to the model indicators, and in particular to the t/F test statistics, in order to obtain robust probability values or heteroskedasticity and autocorrelation consistent covariance estimators [33]. In these models, a backward approach is implemented to sequentially remove variables from the model, starting with the weakest variables that are not statistically significant at the 0.10 level, until a model with significant coefficients is obtained [29]. It is worth noting, however, that the backward approach did not produce a significantly better model than the initial models. Therefore, the initial regression models are reported in Table 6.

Here t is the current value of the month and t-1 is the value of the previous month. The variable ΔP_i represents the amount of change per unit time for the P_i time series with the Δ operator. The models have an autocorrelation problem, the HAC approach is used, and the obtained p-values of the t test coefficients are given in parentheses. Underlined coefficients are statistically significant at 0.05 and italicized coefficients are significant at 0.10.

Table 6. Variable coefficients and probability values of regression models.

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Const	-0.0146 (.265)	0.0109 (.463)	0.0295 (.408)	0.0048 (.755)	-0.0187 (.288)
P_i	-0.3313 (.452)	-0.3342 (.548)	-0.2685 (.677)	0.1096 (.632)	-1.0057 (.001)
PPI	-0.0569 (.337)	0.1062 (.161)	0.1621 (.186)	-0.0727 (.069)	0.0515 (.442)
$IntR$	0.1974 (.156)	-0.1011 (.415)	0.0889 (.662)	0.1612 (.370)	0.0912 (.400)
USD	0.6864 (.150)	-0.5642 (.168)	-1.4303 (.153)	-0.1260 (.796)	0.7342 (.199)
CCI	1.2644 (.168)	0.3939 (.539)	1.0396 (.513)	1.3086 (.151)	-2.2287 (.026)
PMI	-0.6693 (.026)	-0.1184 (.751)	-0.8663 (.045)	-0.4311 (.285)	-0.5308 (.093)
GDP	1.5885 (.117)	0.4023 (.604)	-0.0479 (.974)	-0.4457 (.379)	1.7129 (.005)
$Q_i(t-1)$	-0.4657 (.000)	-0.3380 (.000)	-0.5054 (.000)	-0.3902 (.000)	-0.4536 (.000)
ΔP_i	0.1106 (.682)	-0.1993 (.568)	0.4668 (.233)	-0.0095 (.919)	0.6370 (.000)
R^2	0.2411	0.1591	0.2929	0.1945	0.2661
F test p-value	0.0002	0.0179	0.0000	0.0030	0.0000

According to the OLS results in Table 6, the PMI and especially lagged average product sales ($Q_i(t-1)$) have negative effects on current average product sales. In model 5, product price and price change also have an effect on average product turnover. However, although the F-statistics are quite satisfactory, the predictive power of the models is almost 25% with respect to R^2 when forecasting the average product sales of a month, which is the dependent variable. It can therefore be concluded that it is controversial to build a strong OLS model with the variables/data considered in the present study. Compared to the results of [1], which has a similar OLS application and shows in particular that PMI and P_i have an effect on sales, only 2 models in our study showed similar results in parallel with that study.

5.2. Model estimate results via ANN

The representation of the inputs and the output of the sales forecasting model applied to the 5 products is shown in Figure 3. $PPI, IntR, USD, CCI, PMI, GDP$ are common variables used in the model inputs to determine the amount of product sales. The P_i monthly average product sales (tonnes), the change in the P_i value, and the previous value of the Q_i monthly average product prices (USD) depend on the 5 products. Finally, the forecast of the next Q_i value is used in the model output.

Prior to the model training phase, the missing data were filled using the KNN approach. For example, missing data in the 3rd sales data were filled using this method. The min–max normalization process brings the features used in the model input to the same set of numbers. A neural network model was used to create a sales prediction model. Five different models were built for the data for 5 sales. Nine variables are used as input for 5 models. For the sales prediction model 70% training, 15% validation, and 15% test data were used (95-20-20 samples). The number of neurons in the hidden layer was taken as 50. The training algorithm chosen was Levenberg–Marquardt.

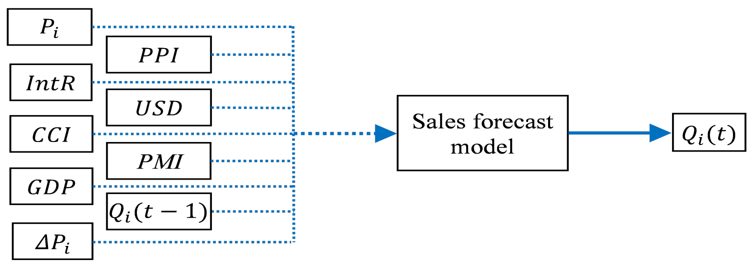

Figure 3. Representation of the inputs and the output of the model applied for 5 products.

Comparing the test data results in Table 7, it can be seen that the performance of models 2, 3, and 4 is higher than that of models 1 and 5. The overall performance of models 2, 3, and 4 remained above 80%. The overall performance of models 1 and 5 is around 60%.

Table 7. Change in R (Pearson correlation coefficient) value as a result of training for 5 models.

	Training	Validation	Test	Overall
Model 1	0.8773	0.2332	0.3387	0.6112
Model 2	0.9585	0.7036	0.7370	0.8971
Model 3	0.9333	0.7862	0.6608	0.8614
Model 4	0.9112	0.5812	0.7999	0.8228
Model 5	0.7292	0.4172	0.4438	0.5990

The performance values for the training, validation and test datasets for model 4 in terms of MSE and R are given in Table 8. In addition, the error distributions of the training, validation, and test data for model 4 are shown in Figure 4. Finally, the regression curve for the training, validation, and test data for model 4 is shown as an example in Figure 5.

Table 8. Performance values of model 4 for training, validation, and test stages.

	Samples	MSE	R
Training	95	6.31681e-3	9111224e-1
Validation	20	2.262262e-2	5.81181e-1
Testing	20	1.16555e-2	7.99955e-1

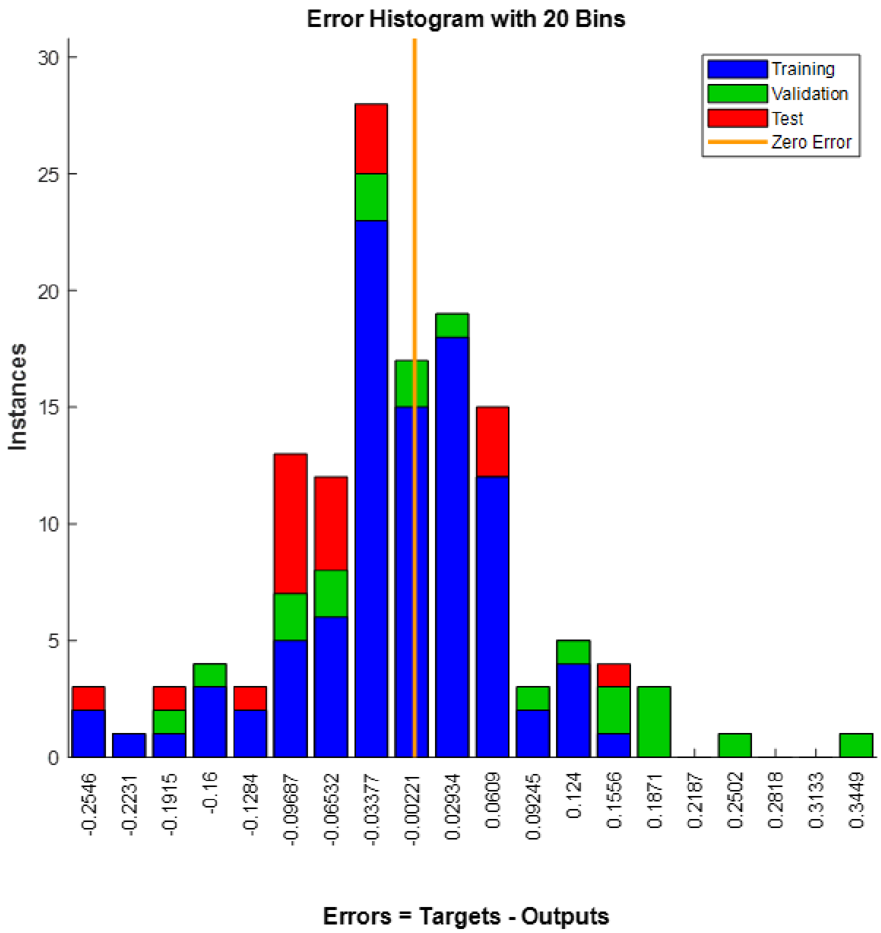

Figure 4. Errors distributions of training, validation, and test data for model 4.

The rank of importance of the features applied in the model inputs with principal component analysis was examined separately for each product. The order of importance of the features is given for each model as shown in Table 9. For model 1, the most important feature is CCI, while $IntR$ is the least important. For model 2, the most important parameter is USD and the least important is PPI. Looking at model 3, $IntR$ is the most important and GDP the least important. For model 4, USD is the most important feature and ΔP_4 is the least important. Finally, while P_5 is the most important feature for model 5, ΔP_5 is the least important feature. The level of effectiveness of all the features based on the model can be examined in Table 9 and visually by examining the spider graph in Figure 6 in more detail.

5.3. Variables' priorities evaluation via GRA using ANN results

The priorities of the variables in relation to the models presented in Table 9 and Figure 6 are clear, but in order to obtain an integrated result from these results obtained by ANN, we used GRA to obtain a more understandable and integrated result for the priorities of the variables. Since each product used in the models addresses different manufacturing industries, it was observed that it is affected by different variables according to the structure of the manufacturing industry. Here general conclusions are drawn and evaluated with GRA.

Table 9. Change in R (Pearson correlation coefficient) value as a result of training for 5 models.

	Model 1		Model 2		Model 3		Model 4		Model 5	
Rank	Variable	Value	Variable	Value	Variable	Value	Variable	Value	Variable	Value
1	CCI	2.96	USD	46.49	$IntR$	35.30	USD	27.34	P_5	4.44
2	$Q_1(t-1)$	2.03	P2	37.08	USD	30.77	P_4	25.60	$IntR$	3.95
3	P_1	1.86	$Q_2(t-1)$	27.10	$Q_3(t-1)$	17.78	PMI	22.68	USD	3.85
4	GDP	1.29	CCI	19.34	CCI	12.55	$IntR$	21.97	$Q_5(t-1)$	2.95
5	ΔP_1	1.19	$IntR$	18.21	P_3	12.22	$Q_4(t-1)$	13.79	PPI	1.77
6	PMI	0.92	PMI	15.46	PMI	8.83	CCI	11.60	PMI	1.51
7	USD	0.26	ΔP_2	3.54	PPI	5.80	PPI	6.40	GDP	0.50
8	PPI	0.23	GDP	1.93	ΔP_3	1.92	GDP	2.23	CCI	0.41
9	$IntR$	0.10	PPI	0.69	GDP	0.04	ΔP_4	2.09	ΔP_5	0.18

Table 10 shows the priorities of the variables normalized on the basis of Table 8 (obtained with the ANN method). In order to compare the results of the priority of these variables obtained by the 5 models, it is useful to use linear normalization for the priority elements. The scoring with normalization was done in order to be able to compare the results of the different model priorities and to avoid an excessive allocation to the relatively high scoring series when calculating the integrated results of the 5 models.

As the ranks in Table 9 show, the variables that affect the level of sales of an iron and steel company can be ordered in terms of their priorities. Therefore, we need to use an integrated calculation procedure to order the variable priorities in this problem. It is accepted that there are other points to be considered in this integrated priority calculation phase. The obtained variable priorities should not be accepted as equal, because models' ANN performance indicators or product sales volume/price etc. may have a affect on these integrated priorities. Table 11 shows the weighting scenarios for the integrated prioritization calculation. Accordingly, the calculation of integrated priorities allows us to use MCDMs. GRA is an MCDM and is derived from gray system theory (developed by Deng in the 1980s), which considers the relationship between the values of a reference series and other series [34]. Since GRA takes into account the similarity between the series being compared, it has some advantages over alternative MCDMs, and it has been used in a variety of studies in recent years. For the computational steps of GRA, see the research of [34]. In order to obtain an integrated result from the variable priority results of the models via ANN, the calculated GRA index scores and their ranks for each variable in terms of scenarios are shown in Table 12. In these integrated results, the test period performance of the models (Scnr2), the general performance of the models (Scnr3), the amount of product sales in relation to the volume of product sales (Scnr4), and the product sales price index (Scnr5) are indicators of the severity of the consideration of the model results. If each priority finding is accepted equally in terms of models, Scnr1 can be taken into account. The GRA index scores in the last column are calculated using the average of the 5 scenario GRA index scores.

According to Table 12, USD has the highest priority and thus has a dominant influence on forecasting the sales amount of this iron and steel company with respect to the ANN method. Product price (P_i) and interest rate ($IntR$) also have quite an overwhelming effect on predicting the sales amount according to all scenarios. Last month's price change (ΔP_i), GDP, and PPI seem to have the weakest effect on predicting the sales amount. These results are quite different from the OLS results. As $Q_i(t-1)$ and PMI are the most important determinants in OLS models, they are moderately important variables in ANN models. As USD

and $IntR$ are the least important determinants in OLS models, they are the most important variables with P_i in ANN models.

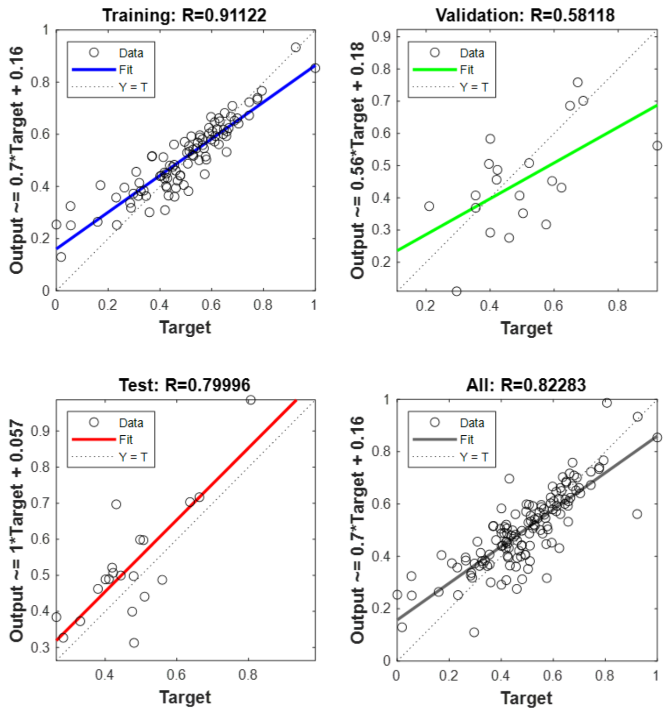

Figure 5. Regression curve for training, validation, and test data for model 4.

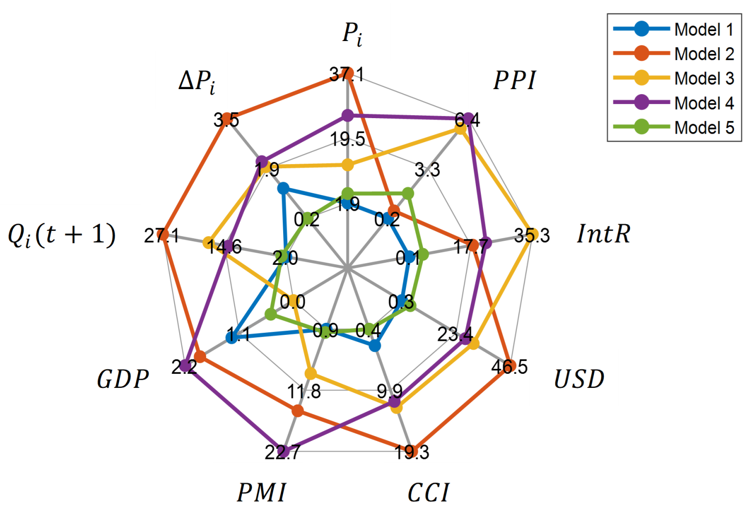

Figure 6. Spider graph of feature importance for all models.

Table 10. The priority ranks of the variables affecting the sales amount of an iron and steel company.

	Normalized priorities for models					Normalized priority ranks for models				
Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
P_i	0.615	0.795	0.345	0.931	1.000	3	2	5	2	1
PPI	0.045	0.000	0.163	0.171	0.373	8	9	7	7	5
$IntR$	0.000	0.383	1.000	0.787	0.885	9	5	1	4	2
USD	0.056	1.000	0.872	1.000	0.862	7	1	2	1	3
CCI	1.000	0.407	0.355	0.377	0.054	1	4	4	6	8
PMI	0.287	0.322	0.249	0.815	0.312	6	6	6	3	6
GDP	0.416	0.027	0.000	0.006	0.075	4	8	9	8	7
$Q_i(t-1)$	0.675	0.577	0.503	0.463	0.650	2	3	3	5	4
ΔP_i	0.381	0.062	0.053	0.000	0.000	5	7	8	9	9

Table 11. The scenarios' weight calculation of each model for priority via GRA (NESO: normalize the elements (variable values) to sum one, dividing each model elements by the sum over all elements. ModPerR: R value of 5 models for model performances in Table 7).

Scenario name and the way of calculating priorities	Calculated weights for models				
	Model 1	Model 2	Model 3	Model 4	Model 5
Scnr1: Each model has equal weight.	0.2000	0.2000	0.2000	0.2000	0.2000
Scnr2: By using the test period success values of ModPerR via NESO.	0.1137	0.2473	0.2217	0.2684	0.1489
Scnr3: By using the overall success values of ModPerR via NESO.	0.1612	0.2366	0.2272	0.2170	0.1580
Scnr4: By using each Qi series averages subject to the model via NESO	0.2100	0.0873	0.0668	0.5602	0.0757
Scnr5: By using the values of each pair multiplication of averages of Qi and Pi series subject to the model via NESO	0.1887	0.0688	0.0641	0.6331	0.0453

Table 12. Variable priority indexes via GRA and their ranks in terms of scenarios.

	Scnr1		Scnr2		Scnr3		Scnr3		Scnr5		Mean of GRA scores	
	GRA	Rank	GRA	Rank	GRA	Rank	GRA	Rank	GRA	Rank	GRA	Rank
P_i	0.717	2	0.720	2	0.706	2	0.785	2	0.778	2	0.741	2
PPI	0.374	7	0.371	7	0.371	7	0.370	7	0.371	7	0.371	7
$IntR$	0.659	3	0.680	3	0.667	3	0.639	3	0.630	3	0.655	3
USD	0.785	1	0.848	1	0.814	1	0.854	1	0.833	1	0.827	1
CCI	0.537	5	0.495	6	0.520	5	0.546	5	0.555	5	0.530	5
PMI	0.478	6	0.499	5	0.483	6	0.614	4	0.591	4	0.533	4
GDP	0.364	8	0.352	9	0.358	8	0.359	8	0.363	8	0.359	8
$Q_i(t-1)$	0.544	4	0.531	4	0.537	4	0.516	6	0.523	6	0.530	6
ΔP_i	0.361	9	0.353	8	0.358	9	0.357	9	0.359	9	0.358	9

5.4. Evaluation and discussion of GRA results

In this part, the overwhelmingly influential variables of USD, P_i, and $IntR$ in predicting the amount of sales are evaluated and compared with the findings of related literature.

The finding of USD as an effective variable has shown parallelism with other studies conducted in Iran [23] and Türkiye [24]. The products and raw materials used by companies in the iron and steel industry are subject to global trade and the trade is usually conducted in USD. Changes in the USD exchange rate may affect the costs, profitability, and prices of producers, which in turn may affect sales volumes. In addition, if fluctuations in Turkish lira prices as a result of changes in exchange rates cause delays in demand and investment decisions, companies may be tempted to purchase iron and steel products ahead of a possible increase in exchange rates in order to hedge themselves.

The $IntR$ is also accepted as an effective variable in the study by [10] for the UK steel industry. It was found to be an indicator of the general economic environment rather than having a direct effect on demand in the sector and to have an indirect effect. In cases in which $IntR$ is high, investment tends to fall as capital tends to earn interest and this can affect the cost of raw materials and therefore the price of the final product in the long run.

According to basic economic theory, it is normal for P_i to affect the quantity sold. However, it was not the most influential variable in our study. In recent years, when inflation in Türkiye has been on an upward trend, especially the export-driven demand has increased and balanced the decline in the domestic market[4]. From another point of view, considering that the iron and steel industry is a sector that supplies raw materials to other industries, it can be assumed that the expectation of an increase in P_i causes the producing companies to stock up on raw materials and increase demand periodically. This situation may be due to the fact that there is less competition compared to the sector to which the product is addressed, that there are different competitors in the sector, and that it is influential for products with low added value.

6. Conclusion

In the present study, 5 different models were developed for different products with the aim of estimating the output value Qi based on input variables including Pi, PPI, $IntR$, USD, CCI, PMI, GDP, $Qi(t-1)$, and ΔPi. These models underwent training, validation, and testing processes. Before entering the model, all variables were subjected to min–max normalization within their respective ranges. This approach ensured that the ANN model was sensitive to changes in all variables, rather than being overly influenced by a single variable. Model validation was performed using the 10-fold cross-validation method. In the training phase, estimation plots and linear regression curves were generated for the output. Test performance was evaluated for the model type that produced the lowest error during training. In general, the test performance values for the ANN models were as follows: models 2, 3, and 4 achieved 0.8971, 0.8614, and 0.8228, respectively, while models 1 and 5 achieved 0.6112 and 0.5990, respectively. An examination of the importance of the variables in the models showed that in models 2, 3, and 4 the variables USD, CCI, and PMI dominated the estimation processes. Conversely, models 1 and 5 performed less well because the variables on which they were based were not sufficiently distinctive. Given the dominance of the USD, CCI, and PMI variables in models 2, 3, and 4 and the high performance of these models, it can be concluded that these variables contributed significantly to their positive performance. Overall, the results suggest that ANN-based models have robust model validation and generalization capabilities even when applied to limited datasets.

In order to consolidate these findings from the ANN models, we used GRA, incorporating assigned weights for each model's results in terms of test period/general performance via ANN, product sales amount, etc. According to the integrated GRA results across all 5 scenarios, it is evident that the USD, Pi, and $IntR$ variables have the most significant influence on determining the sales of an iron and steel company via an ANN. Consequently, iron and steel companies should prioritize USD, Pi, and $IntR$ when forecasting their future sales and planning. This implies that in economies with higher than average inflation rates, nonderivative trading firms may be influenced in terms of iron and steel demand as a hedge against price increases due to exchange rate risk.

[4]TSEA (2022). Turkish Steel Exporters' Association [online] Website http://www.cib.org.tr/tr/istatistikler.html [accessed 03 August 2023]

Furthermore, the present study reveals two significant outcomes. First, ANN models can provide superior performance compared to classical regression models. Using the same dataset, the estimation performance of regression models was significantly lower than that of ANN models. Second, it was concluded that the use of MCDMs such as GRA can be beneficial in obtaining comprehensive results when faced with the challenge of making inferences from numerous ANN results containing identical variables, thereby obtaining integrated insights based on the performance levels of the models.

In future studies, rather than focusing on a single company, data from different large companies operating in different countries can be used to develop hybrid artificial intelligence-based models. These models could be utilized to identify characteristics that affect predictive performance, assign weights to them, and select them based on their degree of importance. This approach has the potential to reduce parameter uncertainty in the model.

References

[1] Pekkaya M, Uysal Z. Pandemi döneminde demir çelik işletmelerinde satış değerlendirmesi ve satış modellemesi. Uluslararası Yönetim İktisat ve İşletme Dergisi 2021; 17: 69-83 (in Turkish).

[2] Altan A, Karasu S, Bekiros S. Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. Chaos, Solitons & Fractals 2019; 126: 325-336.

[3] Chang V. Towards data analysis for weather cloud computing. Knowledge-Based Systems 2017; 127: 29-45.

[4] Chen A, Blue J. Performance analysis of demand planning approaches for aggregating, forecasting and disaggregating interrelated demands. International Journal of Production Economics 2010; 128 (2): 586-602.

[5] Jeon S, Hong B, Chang V. Pattern graph tracking-based stock price prediction using big data. Future Generation Computer Systems 2018; 80: 171-187.

[6] Karasu S, Altan A, Bekiros S, Ahmad W. A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. Energy 2020; 212: 118750.

[7] Gashler MS, Ashmore SC. Modeling time series data with deep Fourier neural networks. Neurocomputing 2016; 188: 3-11.

[8] Huh KS. Steel consumption and economic growth in Korea: long-term and short-term evidence. Resources Policy 2011; 36 (2): 107-113.

[9] Evans M. Modelling steel demand in the UK. Ironmaking & Steelmaking 1996; 23 (1): 17-24.

[10] Abbott AJ, Lawler KA, Armistead C. The UK demand for steel. Applied Economics 1999; 31 (11): 1299-1302.

[11] Crompton P. Future trends in Japanese steel consumption. Resources Policy 2000; 26 (2): 103-114.

[12] Barska M. Demand forecast with business climate index for a steel and iron industry representative. Metody Ilościowe w Badaniach Ekonomicznych 2014; 15 (2): 27-36.

[13] Crompton P, Wu Y. Bayesian vector autoregression forecasts of Chinese steel consumption. Journal of Chinese Economic and Business Studies 2003; 1 (2): 205-219.

[14] Evans M. An alternative approach to estimating the parameters of a generalised Grey Verhulst model: an application to steel intensity of use in the UK. Expert Systems with Applications 2014; 41 (4): 1236-1244.

[15] Chang PC, Wang YW, Tsai CY. Evolving neural network for printed circuit board sales forecasting. Expert Systems with Applications 2005; 29 (1): 83-92.

[16] Psiloglou BE, Giannakopoulos C, Majithia S, Petrakis M. Factors affecting electricity demand in Athens, Greece and London, UK: a comparative assessment. Energy 2009; 34 (11): 1855-1863.

[17] Efendigil T, Önüt S, Kahraman C. A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: a comparative analysis. Expert Systems with Applications 2009; 36 (3): 6697-6707.

[18] Hicham A, Mohammed B, Anas S. Hybrid intelligent system for sale forecasting using Delphi and adaptive fuzzy back-propagation neural networks. Editorial Preface 2012.

[19] Oğcu G, Demirel OF, Zaim S. Forecasting electricity consumption with neural networks and support vector regression. Procedia - Social and Behavioral Sciences 2012; 58: 1576-1585.

[20] Azadeh A, Neshat N, Mardan E, Saberi M. Optimization of steel demand forecasting with complex and uncertain economic inputs by an integrated neural network–fuzzy mathematical programming approach. The International Journal of Advanced Manufacturing Technology 2013; 65: 833-841.

[21] Aziz AA, Mustapha NHN, Ismail R. Factors affecting energy demand in developing countries: a dynamic panel analysis. International Journal of Energy Economics and Policy 2013; 3 (4): 1-6.

[22] Akyurt İZ. Talep tahmininin yapay sinir ağlarıyla modellenmesi: yerli otomobil örneği. Ekonometri ve İstatistik Dergisi 2015; 23: 147-157 (in Turkish).

[23] Koochakpour K, Tarokh MJ. Sales budget forecasting and revision by adaptive network fuzzy base inference system and optimization methods. Journal of Computer & Robotics 2016; 9 (1): 25-38.

[24] Ecemiş O. Model ağaç yöntemiyle satış tahmini: paslanmaz çelik sektöründe bir uygulama. The Journal of Academic Social Science 2019; 84 (84): 336-350 (in Turkish).

[25] Jain AK, Mao J, Mohiuddin KM. Artificial neural networks: a tutorial. Computer 1996; 29 (3): 31-44.

[26] Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Journal of Pharmaceutical and Biomedical Analysis 2000; 22 (5): 717-727.

[27] Pekkaya M, Pulat Ö, Koca H. Evaluation of healthcare service quality via Servqual scale: an application on a hospital. International Journal of Healthcare Management 2019; 12 (4): 340-347.

[28] Pekkaya M, Pulat Ö, Zeydan İ. Service quality determiners in higher education: the student's perspective. International Journal of Services, Economics and Management 2023; 14 (3): 270-300.

[29] Dökmen G, Pekkaya M, Saymaz N. Sigara bağımlılığı ve devletin sigara tüketimi ile mücadele yöntemleri arasındaki ilişki. Maliye Dergisi 2019; 176: 599-623 (in Turkish).

[30] Madhukumar M, Sebastian A, Liang X, Jamil M, Shabbir MNSK. Regression model-based short-term load forecasting for university campus load. IEEE Access 2022; 10: 8891-8905.

[31] Singh U, Rizwan M, Alaraj M, Alsaidan I. A machine learning-based gradient boosting regression approach for wind power production forecasting: a step towards smart grid environments. Energies 2021; 14 (16): 5196.

[32] Pekkaya M. ARFIMA ve FIGARCH yöntemlerinin Markowitz ortalama varyans portföy optimizasyonunda kullanılması: İMKB-30 endeks hisseleri üzerine bir uygulama. İstanbul Üniversitesi İşletme Fakültesi Dergisi 2013; 42 (1): 93-112 (in Turkish).

[33] Hayes AF, Cai L. Using heteroskedasticity-consistent standard error estimators in OLS regression: an introduction and software implementation. Behavior Research Methods 2007; 39: 709-722.

[34] Hamzaçebi C, Pekkaya M. Determining of stock investments with grey relational analysis. Expert Systems with Applications 2011; 38 (8): 9186-9195.