



Citation: Kayakuş, M., Terzioğlu, M., & Yetiz, F. (2022) Forecasting housing prices in Turkey by machine learning methods. *Aestimum* 80: 33-44. doi: 10.36253/aestim-12320

Received: November 19, 2021

Accepted: March 9, 2022

Published: October 17, 2022

Copyright: ©2022 Kayakuş, M., Terzioğlu, M., & Yetiz, F. This is an open access, peer-reviewed article published by Firenze University Press (<http://www.fupress.com/ceset>) and distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

ORCID:

MK: 0000-0003-0394-5862
MT: 0000-0002-4614-7185
FY: 0000-0001-5480-9268

Original Articles - Urban, land, environmental appraisal and economics

Forecasting housing prices in Turkey by machine learning methods

MEHMET KAYAKUŞ^{1,*}, MUSTAFA TERZIOĞLU², FILİZ YETİZ³

¹ Department of Management Information Systems, Akdeniz University, Turkey

² Department of Accounting and Tax Department, Akdeniz University, Turkey

³ Department of Banking and Finance, Akdeniz University, Turkey

E-mail: mehmetkayakus@akdeniz.edu.tr, mterzioglu@akdeniz.edu.tr, filizyetiz@akdeniz.edu.tr

*Corresponding author

Abstract. In this study, decision tree regression, artificial neural networks (ANN) and support vector machines (SVM) methods are applied by using monthly data for the period 2013-2020 in the estimation of housing sales in Turkey. In the analysis, the volume of individual mortgage loans offered by banks, the average annual interest rate of mortgage loans from macroeconomic and market variables, the consumer price index (CPI), the BIST 100 index, the benchmark bond interest rate, gold prices and the values of the US dollar and Euro Turkish lira and the housing sales price per square meter in Turkey are used. As a result of the analysis carried out on the model created house sales prices in the Turkish housing market have been successfully estimated and in the light of these estimates, it is determined that banks can guide banks in the creation of various credit packages and appropriate loan targets to support the housing sector.

Keywords: Home sales price prediction, Decision tree regression, Artificial neural networks, Support vector machines.

JEL codes: O18, R33.

1. INTRODUCTION

The need for housing, as well as a durable consumer good and an investment vehicle, is an important feature that distinguishes the housing market from other markets (Lacoviello, 2000). The revival in housing sales in the housing market will positively affect the development of the construction sector. Mobility in the construction sector will have a domino effect and affect its sub-sectors and contribute to the economic growth of countries. Therefore, it is important that demand and supply changes in the housing market are accurately predicted. In terms of financing, forecasts for demand and supply in the housing market will guide the banking sector in establishing credit and product targets to support the housing market. By taking advantage of these forecasts, banks will be able to help the housing market grow steadily by creating new low-cost and affordable

product packages and credit campaigns for the needs of the housing market.

In addition, the housing market has been affected in different dimensions in the COVID-19 global pandemic that has affected the whole World. In line with the restrictions experienced during the epidemic process, it is seen that households who have to spend more time in housing have started to prefer detached garden houses and this trend is thought to increase in the future. This will lead to increases in demand for different types of housing. In addition, it is foreseen that the supply-side problems experienced during the epidemic process will continue. Inflationary pressures, especially in countries, and the resulting cost and interest rate increases on mortgage loans are among the significant problems of housing supply. These changes expected after the epidemic in demand and supply make it inevitable to make demand predictions using current methods for housing markets in developing countries, especially in Turkey.

The housing market is both an important indicator of the economy and an important indicator of consumer spending and total prosperity (Greenwood and Hercowitz, 1991). The housing sector has a close relationship with other sectors and the vitality of the sector drives other sectors (Öztürk and Fitöz, 2009, p.23). For these reasons, it is important to estimate the housing market. In the literature, Abidoye and Chan (2017), Mohd Radzi et al. (2012), Nguyen and Cripps (2001), and Zainun et al. (2010) used artificial neural networks in their studies of the housing market. One of the objectives of our study is to evaluate the results of decision tree regulation, support vector machines and artificial neural networks methods in the forecast of the housing market in Turkey. In addition, it is to emphasize the importance of incentives and policies to improve this sector due to the importance of the housing market in the economy. Another goal is to guide the financial and banking sector in the development of new products and services for the housing market.

In the study, a preliminary estimate of housing sales prices was made by using the importance of the housing market in the economy, bank mortgage loan volume and some macroeconomic and market indicators as variables. In the study, decision tree regression, artificial neural networks and support vector machines methods and the estimation of housing sales prices were investigated using Turkey's monthly data for the period 2013-2020. In the data set of the model, consumer price index (CPI), volume of individual mortgage loans of banks, average annual interest rate of mortgage loans, benchmark bond interest rate, gold prices, Euro and US dollar Turkish lira

value, BIST 100 index and average Turkish housing sales price per square meter variables were preferred. The number of scientific studies that emphasize the importance of the housing market in terms of economic structure and research the structural characteristics of the housing market for the Turkish economy is limited in the literature.

When some of the limited studies were examined, Özkan et al. (2007) used artificial neural networks and regression methods for housing price estimation in Selçuk district of Konya. In Selim (2009) study, he used hedonic regression and artificial neural networks methods to estimate house prices using 2004 household budget survey data for Turkey. Temür et al. (2019) estimated housing sales in Turkey using 124 months of data for the period 2008-2018.

In the study, the conceptual framework related to the importance of housing finance and the financial support provided by banks to the housing market will be examined, and research on housing market sales and sales price will be included in the literature section. In the ongoing sections of the study, the methods and methods used in the forecasting of housing sales will be examined and evaluations and recommendations will be made in the result section. Another purpose of this study is to support the policies of regulatory and supportive institutions for the housing market thanks to this model. In this way, these institutions will be able to perform actions that will bring households together with more predictable house prices.

2. HOUSING FINANCE AND ITS PLACE IN THE ECONOMY

Housing finance is the way to direct those who need funds to obtain housing from those who have surplus funds, i.e. from the resources of savers. In other words, it is the meeting of those who offer funds for housing purchase and those who demand funds (Kömürlü and Önel, 2007). A good housing finance system created in accordance with the needs is expected to both increase housing supply and improve housing quality (Berberoglu and Teker, 2011).

It has come to the fore that the costs of housing investments are quite high and that people who want or demand housing needs do not have the savings to cover this cost with their income at a time, and that there should be a form of financing that they can easily finance by spreading their payments over the long term. In this context, mortgage loans that are used by showing the housing guarantee requested in exchange for loans

and establishing mortgages on them form the basis of housing finance (Ayan, 2011).

The indirect or direct effects of the housing market on the economy are realized by housing investments. Housing investments directly affect the economy by increasing employment opportunities, while increased demand for construction materials and other durable consumer goods indirectly affects the economy. Housing investments in the housing market contribute to economic growth through their impact on employment, labor productivity, total savings and investments (Harris and Arku, 2006). Residential manufacturing in the construction sector and other related sectors are accelerating economic growth.

3. LITERATURE RESEARCH

The factors in the formation of house prices and the studies aimed at estimation are mainly divided according to the use of micro and macro variables.

In studies using micro variables, the properties of the housing unit are used. These include variables such as the number of rooms (such as the number of halls and bathrooms), space it covers, parking lot it has, the gym, spa, garden and children's entertainment areas, etc. In addition, variables such as the distance of the dwelling to the city center, the characteristics of the region you are in, the proximity to the metro, etc. are also used. Abidoye and Chan (2016), Abidoye and Chan (2017), Deng et al. (2018), Ecer (2014), Jayasekare et al. (2019), Pai and Wang (2020), and Yilmazel et al. (2018) have tried to estimate housing prices using the microvariability listed above.

The impact of the above-mentioned micro variables in the forecast of house prices, as well as the economic fluctuations experienced by countries, i.e., macroeconomic variables and changes in market indicators, on house prices is undiscussable. In our study, we wanted to test the ability of these macro variables to predict house prices. For this purpose, macro variables, which are subject to studies as a second aspect in the literature, should predict housing sales and prices and studies on the effect of these variables on house prices are given in detail below.

Cho (1996), Mankiw and Weil (1989), Painter and Redfearn (2002), Poterba (1991), and Rapach and Strauss (2007) investigated the effectiveness of the housing market using some variables showing housing market and macroeconomic sizes in their work. In these studies, the importance of demographic factors, consumer price index, real income, borrowing restriction, unemploy-

ment rate, construction costs, etc. were emphasized in explaining the mobility in house prices.

In Painter and Redfearn (2002), which examined the effects of interest rates on housing rates and housing starts, they investigated U.S. housing markets. Some demographic and economic variables selected by house prices, interest rate, income were used in the data set. Peer integration tests and Vector Error Correction model were analysed with quarterly data from 1965-1999. According to the results obtained, there was a negative-directional relationship between housing rates and housing starts in the short term, while the effects of interest rates on housing starts decreased in the long term.

Wong et al. (2003) in his study for Hong Kong examined the effects of interest rates on house prices for both inflationist and deflationist periods. As a result of the study using the granger causality method, a positive relationship was found between interest rates and house prices in the deflationary period, while it was found that there was a negative-directional relationship in the inflationary period.

In his study, Leung (2004) emphasized a strong correlation between macroeconomic factors and the housing market. He stated that the cycles experienced in the housing market in many years will change the macroeconomic structure of the countries. He stated that globalization and financial integration and housing market performance, especially in emerging economies, will shape the housing market finance and capital markets of these economies and this should be investigated for years to come.

Goodhart and Hofmann (2008) analysed the relationships between loan and house prices and monetary variables. Panel data analysis method was used in the study by using the quarterly data of 17 industrialized countries for 1976-2006. It has been determined that there is a positive and strong relationship between housing prices and mortgage loan and growth in monetary variables. A bi-directional Granger causality relationship was found between house prices and other variables.

Priemus (2010) investigated the effects of the mortgaged housing market crisis on the credit markets and housing market in 2008. In the study, the data of the Dutch housing market for the years 1998-2009 were examined. It was concluded that the Dutch housing market was significantly affected by the crisis, housing construction activities decreased significantly during the crisis, and even housing prices decreased significantly.

Duan et al. (2018) used subject prices as dependent variables, and macroeconomic variables such as personal disposable income, real interest rate, unemployment rate, mortgage loan volume, current account balance and

housing investment amount as independent variables in their study using dynamic spatial panel data method. They stated that the degree of change in macro variables not only determines the degree and direction of market real estate price movements, but also has the ability to affect the global housing market balance.

Temür et al., (2019) estimated the house sales in Turkey with 124 monthly data for the period of 2008-2018 in their study. In the study, by a hybrid model consisting of ARIMA (Autoregressive integrated moving average) and LSTM (long short-term memory) methods, it was concluded that the housing sales in the housing sector were estimated to be as close as possible to the real value.

In the studies conducted in the field of the housing sector with machine learning in the literature, it is seen that these methods are successful in estimating house prices close to the truth. These studies are listed below.

Nguyen and Cripps (2001) examined the estimation of house sales using micro variables by comparing the method of artificial neural networks and multiple regression analysis. Sales prices, area of the apartment, age of the building, number of rooms and bathrooms were used in the study. In the present study, it has been determined that the artificial neural network method is more successful if the correct data are selected.

Limsombunchai and Samarasinghe (2004) used an ANN model to predict house prices in a city in New Zealand. As a result, they reached an estimation result close to the real value.

Khalafallah (2008) predicted real estate sales with artificial neural networks using macro variables. The variables used in prediction; interest rate, time, change of sales over years, average sales period, change in unit value of sales compared to the previous year and transaction volume. As a result of the study, it reached the real value of sales with a tolerance of $\pm 2\%$ with the method of artificial neural networks.

Li et al. (2009) tested the strength of the support vector regression method in predicting property prices in China. To this end, he compared the home price estimation performance of the support vector regression method with the performance of the back propagation neural network method using the macro variables of disposable income, real estate development investment amount, consumer price index, loan interest rates with quarterly data covering the period 1998-2008. They found that the support vector regression method performed better in estimating the house price.

Ghodsi et al. (2010), in their study on the estimation of Iranian housing prices, tested economic variables including income from oil, housing price index, general

price index, cost of construction materials and gross domestic product (GDP) using artificial neural networks architecture. According to the test results, back propagation artificial neural network technique (MAPE- Mean Absolute Percent Error) 0.11698 has been formed. They considered successful the estimation of the relevant macro variables on subject prices.

Ecer (2014) made a house price estimation by using the data of 610 houses belonging to 2013 in Karşıyaka district of İzmir. He used the hedonic model and the MLP model, one of the artificial neural network models, in estimating the house price. The variables used in the model are 83 micro variables belonging to residences. When the results of the two models are compared, it is the result that artificial neural networks predict more successfully than the hedonic model. In the hedonic model, the size of the house, dressing room, shower cabin in the residence, en-suite bathroom, jacuzzi, cellar, built-in wardrobe, blinds, built-in kitchen, water heater, indoor garage, elevator, outdoor swimming pool, proximity to primary school, proximity to the pier and sea view were found to be the most important factors affecting house prices.

Demirel et al. (2016) conducted an application in Talas and Kocasinan municipalities in their studies for the province of Kayseri, and it was stated that artificial neural networks achieved very good results in the valuation of flat-type houses.

Using multiple linear regression, Grum and Govekar (2016) analyzed a model that uses macro variables to demonstrate the relationship between housing prices and macroeconomic indicators in different socio-culturally diverse regions in Sloveyna, Greece, France, Poland and Norway. They used the observable stock index, unemployment rate and industrial production index in these countries as macroeconomic indicators. As a result, they stated that there is a high correlation between macroeconomic indicators and house prices, especially in the unemployment rate and housing prices.

Li and Chu (2017) tried to predict the real estate price change in Taipei with macro variables by using back propagation and radial basis function neural network methods. In the period between 2005 and 2015, macroeconomic indicators such as gross domestic product, m^2 money supply, housing price index, gross national product, growth rate, housing prices / gross national product ratio, consumer price index, new housing loan volume They used it in the estimation of change. They concluded that the relevant macroeconomic variables were not sufficiently successful in predicting housing prices.

Abidoye and Chan (2017) used the artificial neural

networks method in property valuation in Lagos metropolitan real estate market. In the study, real estate sales transactions data (11 independent variables and property values) were obtained from real estate companies operating in Lagos, Nigeria. 370 data and 11 independent variables were used to estimate the house price in Nigeria. As a result, it is concluded that ANN models perform well in predicting house prices accurately and are suitable and reliable for property valuation.

In Yilmazel et al. (2018) artificial neural networks method was used in estimating house prices for sale in Eskişehir housing market in Turkey. Many different physical characteristics such as the size structure of the houses, the number of rooms, whether they are on the first floor, the total number of floors in the residence, whether there is central heating, the number of bathrooms and elevators, built-in kitchen, parking and fiber internet connections, the neighbourhood where the residence is located and the tramway Models with distance variables were established. As a result of the study, it was determined that the ANN method is an effective tool in estimating house prices.

Pai and Wang (2020) tried to satisfy Taiwan housing prices using the methods of least squares vector regression, decision tree, general regression neural networks and back propagation neural networks. In order to estimate house prices, 23 variables (number of rooms, distance to the city, park width, etc.) that determine the basic characteristics and location of housing types were used as independent variables. They concluded that the predictive power of the machine learning methods used in the results is applicable to house prices. However, these methods are quite successful in terms of the results of least-squares vector regression compared to other methods.

Ghodsi et al. (2010) used economic variables including income from oil, general price index, housing price index, gross domestic product (GDP) and cost of construction materials as independent variables in their study on the estimation of Iranian housing prices. They used artificial neural network architecture as a method. According to the test results, back propagation artificial neural network technique (MAPE-Mean Absolute Percentage Error) was 0.11698. They found the estimation on the subject prices of the relevant variables successful.

4. MATERIAL AND METHOD

In the study, artificial neural networks, decision trees regression method and support vector regression were used from machine learning methods to estimate

the price of housing m² in Turkey. Statistical analysis methods were used to determine which of these methods was more successful.

4.1 Data Set

In this study, three different machine learning methods were used to estimate the price of residential m². These are artificial neural networks, decision trees regression method and support vector regression methods. In the models designed, 9 variables were determined as inputs to estimate the price of house square meters (m²) in Turkey with output value. The study used the following data for 95 months between January 2013 and November 2020 as input variables:

- Central Bank of the Republic of Turkey (CBRT) House Price Index
- Mortgage Loan Average Annual Interest Rate
- Dollar/TL
- Euro/TL
- BIST 100 Index
- Consumer price index (CPI) (2003=100) (TURK-STAT) (Monthly)
- 2-Year Government Bond Indicator Interest (%)
- Gold (\$1 Troy Ounces)
- Individual Mortgage Loans (Thousand TL)

4.2 Support Vector Regression

It is a machine learning algorithm based on supervised learning theory that uses statistical calculation methods recommended by Vapnik for the solutions of classification and regression problems (Vapnik, 1999). Support vector machines (SVM) are a type of machine learning tool which can solve classification, regression and innovation detection problems with better generalization compared to other learning algorithms. The type of SVM used in regression applications is called SVR (Support Vector Regression), and the type used in classification applications is called SVC (Support Vector Classification).

The principle of operation of the SVM is based on the estimation of the most appropriate decision function that can distinguish between the two classes, in other words, the identification of hyper-plane, which can most appropriately distinguish between the two classes (Kavzoglu and Çölkesen, 2010). SVM is divided into two according to the linear detachment and inability to separate the data set (Vapnik, 1999).

Linear support is called the optimal hyperplane in vector machines, which linearly divides the data set into

two. Although many decisions can be drawn correctly, the most important decision is to determine the right one.

If the training data consisting of a number of examples for the training of SVM in a linearly detachable two-class classification problem is considered $\{x_i, y_i\}_{i=1, \dots, k}$, the optimal hyper-plane inequalities are as follows:

$$\begin{cases} w \cdot x_i + b \geq +1 & y = +1 \\ w \cdot x_i + b \leq -1 & y = -1 \end{cases} \quad (1)$$

Here $x \in \mathbb{R}^N$ shows an N-dimensional space, w weight vector (normal of hyper-plane), $y \in \{-1, +1\}$ class labels and b trend value (Osuna, 1998). In order to determine the optimum hyper-plane, it is required to determine the two hyper-planes that will form the parallel and boundaries of this plane. The points that make up these hyper-plane are called support vectors, which are expressed as $w \cdot x_i + b = \pm 1$. Figure 1 show the linear support vector machine (Adar and Kilic Deli, 2019).

Nonlinear SVM are algorithms used if the data set cannot be separated by a linear function with a full or specific error. In real-life problems, it is often not possible to separate a data set linearly with the hyper-plane. In this case, the problem caused by the fact that some of the training data remain on the other side of the optimal hyper-plane is solved by the identification of a positive artificial variable (ξ_i). The balance between maxing out the limit and making misclassification errors minimum can be controlled by defining an editing parameter ($0 < C < \infty$) indicated by the positive values area and C. Non-linear optimization problem here (Kavzoğlu and Çölkesen, 2010):

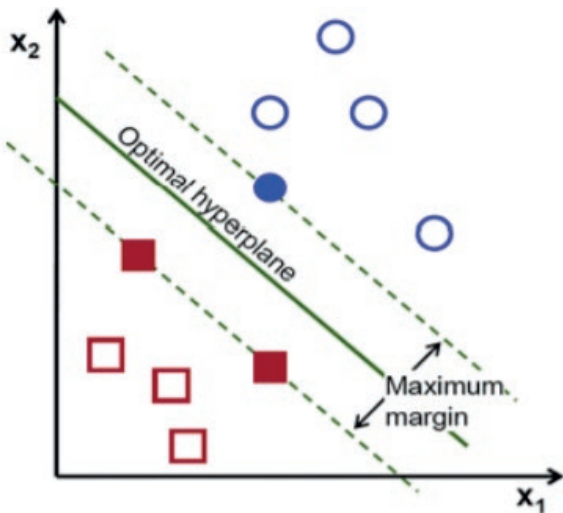


Figure 1. Linear support vector machine (Adar and Kilic Deli, 2019).

$$\min \left[\frac{\|w\|^2}{2} + C \cdot \sum_{i=1}^r \xi_i \right] \quad (2)$$

Depending on the limitations:

$$y_i (w \cdot \varphi(x_i) + b) - 1 \geq 1 - \xi_i \quad \xi_i \geq 0 \text{ ve } i = 1, \dots, N \quad (3)$$

Support vector machines can make nonlinear transformations with the help of kernel function and in this way allow for high-size linear separation of data. Data that cannot be separated linearly in input space is displayed in a high-dimensional space defined as a property space. Thus, the data can be distinguished linearly and the hyper-plane between the classes can be determined (Kavzoğlu and Çölkesen, 2010).

It is essential to determine the kernel function to be used for a classification process with support vector machines (SVM) and the optimum parameters for this function. The most used polynomial function in the literature is radial-based function and normalized polynomial kernel functions.

4.3 Decision Trees Regression

The decision tree method is one of the important machine learning techniques used in forecasting and classification. In this method, it is a model that examines the relationships of arguments with each other and the dependent variable in the form of trees. In the tree model, decision-making points are called nodes. In the tree model, the starting node, which contains all the relationships between variables and is the most complex node, is called the root and begins with the tree structure. According to the relationship between the arguments, each time with a binary branching, the heterogeneous in the plot is divided into sub-nodes that are homogeneous in another (Takma et al., 2017).

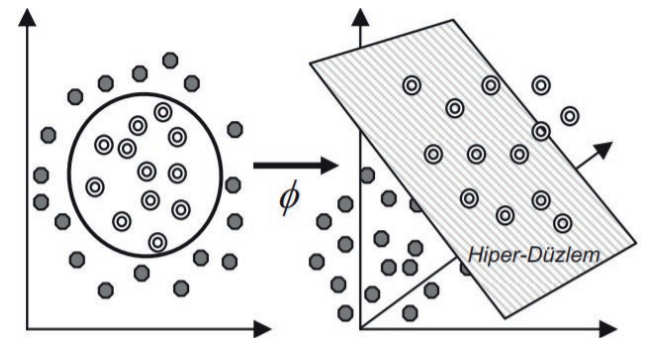


Figure 2. Conversion of data to a higher size with kernel function (Kavzoğlu and Çölkesen, 2010).

The decision tree ranges arguments according to information gain. When a value is asked to be learned from this range during the prediction, the average of the values in the range it learns during the training is found. That is why decision tree regression is cut, not continuous, like other regression models.

The decision consists of two basic stages: splitting and pruning. Splitting is a recurring stage that allows data in the data set to be divided into smaller sub-groups. The first phase begins with the root node containing the entire dataset. In the following stages, the nodes containing subsets of the data are processed. Variables are analysed and the best splitting process is selected after each split operation. After a tree is created, deforestation is used to find unwanted sub-trees or nodes. With the budding process, these can be removed, and the decision tree can be expressed in a more general way. The stop criterion includes several stop rules, tree-building algorithms. These rules are usually based on several factors, such as maximum tree depth, the minimum number of elements handled for partitioning on a node, and the minimum number of elements that should be in a new node (Shooting and Karasoy, 2013).

Decision Trees try to maximize information gain by making choices that reduce the entropy value of the current situation. To do this, it recalculates the error function on each node and selects the state with the lowest error. There are many decision tree algorithms developed to create the decision tree structure. Decision tree algorithms typically expected to create the most appropriate decision tree structure that minimizes generalization error (Aytuğ, 2015). One of the approaches used to create a decision tree is the use of criteria for node allocation. Chaid, Cart, Mars, Quest, Sliq, sprint, ID3, C4.5, and C5.0 are among the node allocation criteria used.

CART is a non-parametric statistical method that uses decision trees to solve classification and regression problems using both categorical and continuous variables. If the dependent variable is categorical, the method is called classification tree and continuously regression trees (Deconinck et al., 2005). The CART algorithm has a structure that creates binary decision trees by separating the relevant set into two subsets that are more homogeneous than another at each stage. The best argument is selected using impurity and variability in change measures (Gini, Twoing, smallest squares deviation). Here, the goal is to produce the most homogeneous sub-groups of data possible for the target variable.

The CART algorithm, in which each node is divided into two at each stage, uses the Gini integrate from these impurity criteria developed to select the best division, such as Gini and Twoing, in determining each division

point (Deconinck et al., 2005). $Gini_{left}$ and $Gini_{right}$ values are calculated for splits on the left and right sides in each attribute (Güner, 2014):

$$Gini_{left} = 1 - \sum_{i=1}^k \left(\frac{L_i}{T_{left}} \right)^2 \quad (4)$$

$$Gini_{right} = 1 - \sum_{i=1}^k \left(\frac{R_i}{T_{right}} \right)^2 \quad (5)$$

where k shows the number of classes, T shows the number of instances on a node, T_{left} shows the number of instances on the left node, T_{right} shows the number of instances on the right node, L_i shows the number of instances in the i category on the left node, R_i shows the number of instances in the i category on the right node.

The Gini index value is calculated by the following formula, n is the number of rows in the training data for each j attribute.

$$Gini_j = \frac{1}{n} (T_{left} \times Gini_{left} + T_{right} \times Gini_{right}) \quad (6)$$

4.4 Artificial neural networks

Artificial neural networks are inspired by the functioning of the biological neuron structure of organisms. Artificial neural networks (ANN) are a modelling tool used to solve complex problems that are accepted by many disciplines. Artificial Neural networks consist of a series of interconnected parallel structures known as neurons or nodes.

An artificial neural network consists of three basic functions: input, hidden and output layer. The input layer (x_1, x_2, \dots, x_n) transmits the information they receive from the external environment to the nervous system. Input values taken into the artificial neural network model are multiplied by coefficients called weight (w_1, w_2, \dots, w_n). Weights are coefficients that determine the effect of inputs through the nervous system, and appropriate values are available through trial and error. Entries are then multiplied by the weights to which they belong, and the threshold value is added to the result. The result is passed through the activation function and transmitted to the output. If the signal is below the threshold value, the output is not produced; output is produced on it. The output layer is the layer where the result is sent to the outside world (Diamond, 2018). Artificial neural network function:

$$NET = \sum_{i=1}^n w_{ij}x_n + \theta_j \quad (7)$$

where n is the number of inputs, x is the input neuron of the network, w_{ij} is the weight of the connection between the neuron, θ_j represents the cell's threshold value and y the output value. The output value of the network is obtained by going through the activation function of the resulting value (y), as shown in Equation 8.

$$y=f(NET) \quad (8)$$

Activation function has linear and nonlinear forms: linear, step, sigmoid and tangent hyperbolic. When selecting the activation function, it is important to make the derivative easily calculable.

Artificial neural networks have two network architectures: forward-feed and feedback networks according to the direction of connections and flow within the network. Forward-feed networks consist of three layers: input, output, and at least one hidden layer. The number of nodes in the hidden layer and hidden layer varies according to the structure of the network. Each node in the input layer depends on each node in the hidden layer. The output values of each layer are called forward feeds to move toward the output so that the input value of the next layer is. The purpose of feedback networks is to optimize the error value by optimizing the weights. Input data is spread over the network to find out an estimate of the output. Weights are systematically updated according to the error information in the estimate. The network is trained by changing weights until the error between training data outputs and the network's predicted outputs is small enough (Hancke and Malan, 1998).

4.5 Evaluation Methods

Coefficient of determination (R^2), Mean Squared Error (MSE), Root Mean Square Error ($RMSE$), Mean Absolute Error (MAE), Mean Absolute Percentage Error ($MAPE$) methods were used to measure the success of predictions of the designed artificial neural network model. The equations for these methods are given below.

$$R^2 = 1 - \frac{\sum (y_i - x_i)^2}{\sum (y_i - y_{ave})^2} \quad (9)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2 \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - x_i}{Y_i} \right|^2} \quad (11)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (12)$$

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{y_i - x_i}{y_i} \right|}{n} \cdot 100 \quad (13)$$

where i shows the number of data, the actual value of y , the estimated value x . According to these criteria, high R^2 and lowest $RMSE$, MSE , MAE and $MAPE$ values determine the most successful model.

5. RESULTS AND DISCUSSION

In the study, it was tried to estimate the price of residential m^2 in Turkey using machine learning methods. In the study, the normalization method was used to improve the performance of machine learning methods and increase accuracy. Since it gave the best result in this study, the original data was normalized with linear transformation in the range of 0-1 using the min-max normalization method from normalization methods. The min-max normalization formula is given in Equation 14.

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (14)$$

x' shows normalized data, x_i input value, the smallest number in x_{min} input set, the biggest number in x_{max} input set.

The nonlinear support vector regression method was used in the study. In the SVR method, tests were performed using Polynomial, Hyper Tangent and Radial basis function (RBF) techniques from core functions; Radial basis function (RBF) was preferred as the core function because it gave the successful result. 80% training of the data set in the study; 20% of it was used as test data. The overlapping penalty is a useful parameter used if input data is not detachable. It determines how much punishment will be given to each incorrectly predicted point, thus increasing the accuracy rate. In this study, overlapping penalty values were tested and the best value was found to give at 10.

In the decision trees regression method, the "Gini Index" was used as the quality measure in which the partition was calculated. Each node has at least 2 minimum records. If the number of records is less than or equal to this number, the tree is no more grown. The tree stores 10,000 records for appearance. The study used 8 threads and therefore the processor or core. This has improved the performance of the study.

Table 1. Test results.

Machine learning methods	R ²	MSE	RMSE	MAE	MAPE
Support vector regression	0,987	0,001	0,027	0,020	0,095
Decision trees regression	0,989	0,001	0,026	0,020	0,066
Artificial neural networks	0,981	0,001	0,033	0,020	0,163

It is modelled to have 100 iterations in the artificial neural network, 4 hidden layers and 4 neurons per hidden layer. In the model, it used a back propagation function in which weight values are updated according to the behaviour of the error function. The values that the parameters used to create models will receive have been tested by giving different values to show the highest performance for each dependent variable, and ideal values have been selected.

The coefficient of determination, R^2 , which indicates how well the data fit into a linear curve, is 1, indicating that the test data provides a linear curve. R^2 value according to SVR method in study 0.987; 0.989 according to the decision tree method and 0.981 according to artificial neural networks method. According to this conclusion, 98.7% of the change in the output variable according to the SVR method, 98.9% according to the decision tree and 98.1% according to artificial neural networks can be explained by input changes. These three methods have been sorting decision trees, SVR and artificial neural networks compared to the success of the determination coefficient.

In the study, *MSE* values were found to be ideal values because they were very close to zero with 0.001. The *RMSE* value is requested to be close to zero; In this study, it was seen that it is in the range of 0.026-0.033

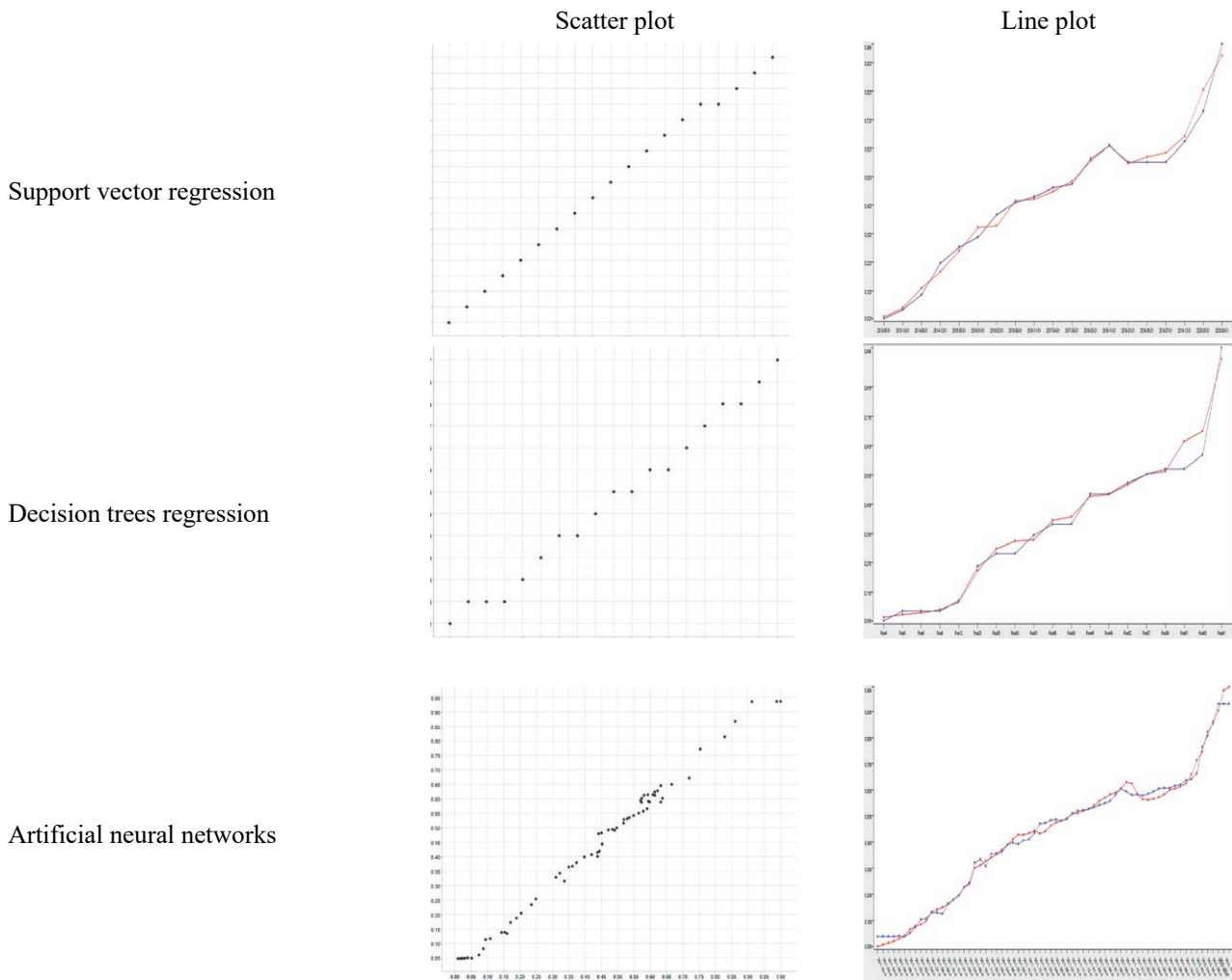


Figure 3. Graphical analysis of test data.

and is very close to the desired value. The three methods were optimal sorting decision trees, SVR and artificial neural networks compared to RMSE value.

In the study, MAE value was 0.020 with an ideal value of zero in three methods. The desired value for MAPE is the smallest value near zero. Accordingly, sorting decision trees have been in the form of regression, SVR and artificial neural networks.

The scatter plot and line plot curves were used to analyze test data in the study. It is seen as comparative in Figure 3.

The scatter plot curve is used to determine the relationship between two different variables. According to the scatter plot curve, which shows the relationship between the actual house m² price and the estimated house m² prices, there is a linear and strong relationship according to all three machine learning methods.

The line plot is a type of chart used to show and compare quantitative values over a period. It is used here to compare actual values with predicted values. When the values shown by the charts are examined, it is seen that all three machine learning methods are successful.

6. Conclusion

Creating solutions to problems in the housing market in developing countries, improving housing construction, diversifying the financing dimension of housing production are supported by the economic policies developed. Studies on housing sales price forecasting are important for creating incentives for the housing market and supporting economic policies. In developing countries, this is necessary for accelerating the growth momentum of countries. Successful forecasts for the housing market will allow banks to provide the funds they need in a shorter time, taking into account their ability to easily reach and pay for their customers who want to buy housing. Thus, the transfer of resources from the surplus of funds to the segment in need of funds will be fluting efficiently. An optimal housing financing system will emerge, and both household housing asses will be facilitated, and the growth performance of countries will improve.

The results obtained are also an indication that our study will make a significant contribution to the literature. It is thought that the studies to be carried out with machine learning for the housing market in Turkey, which is a developing country, can guide the construction of future plans and solutions to problems for the housing market and related sub-sectors, as well as the creation of various financing products suitable for the housing market in the banking system.

REFERENCES

- Abidoye, R. B., & Chan, A. P. (2017). Critical determinants of residential property value: professionals' perspective. *Journal of Facilities Management*, 14(3), 283–300.
- Adar, T., & Kılıç Delice, E. (2019). A literature review on the use of Machine Learning algorithms in health. Paper presented at the 4th International Energy & Engineering Congress, October 24-25, 2019, Gaziantep, Turkey.
- Ayan, E. (2011). Housing finance system and analysis of the applications of housing development administration of Turkey. *The Journal of Accounting and Finance*, (51), 139–156.
- Aytuğ, O. (2015). Comparative performance analysis of Decision Tree algorithms in the corporate bankruptcy prediction. *Journal of Information Technologies*, 8(1), 9–19.
- Berberoğlu, M. G., & Teker, S. (2005). Housing finance and the proposal of an appropriate model for Turkey. *İtüdergisi/b*, 2(1), 58–68.
- Cho, M. (1996). House price dynamics: a survey of theoretical and empirical issues. *Journal of Housing Research*, 7(2), 145–172.
- Çekim, H. Ö., & Karasoy, D. (2013). Cox hybrid model with decision tree and an application on data of tires. *Journal of Statisticians: Statistics and Actuarial Sciences*, 6(1), 41–50.
- Deconinck, E., Hancock, T., Coomans, D., Massart, D., & Vander Heyden, Y. (2005). Classification of drugs in absorption classes using the classification and regression trees (CART) methodology. *Journal of Pharmaceutical and Biomedical Analysis*, 39(1-2), 91–103.
- Demirel, S. B., Reyhan, O., Atasever, Ü. H., & Hayri, M. (2016). The Using of artificial neural networks in flat type real estate valuation, 6. Remote Sensing-CBS Symposium, 5-7 October, 2016, Adana, Turkey, 98–104.
- Deng, Y., Girardin, E., & Joyeux, R. (2018). Fundamentals and the volatility of real estate prices in China: a sequential modelling strategy. *China Economic Review*, 48, 205–222.
- Duan, K., Mishra, T., & Parhi, M. (2018). Space matters: Understanding the real effects of macroeconomic variations in cross-country housing price movements. *Economics Letters*, 163, 130–135.
- Ecer, F. (2014). Comparison of hedonic regression method and Artificial Neural Networks to predict housing prices in Turkey. International Conference On Eurasian Economies, 1-2 July, Skopje, Macedonia.
- Elmas, Ç. (2018). *Artificial intelligence applications*. Ankara, Seçkin Publishing.

- Fidan, U., Uzunhisarcıklı, E., & Çalığışu, İ. (2019). Classification of dermatological data with self organizing maps and support vector machine. *Afyon Kocatepe University Journal of Science and Engineering*, 19(3), 894–901.
- Ghodsi, R., Boostani, A., & Faghihi, F. (2010). Estimation of housing prices by fuzzy regression and artificial neural network. Paper presented at the 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation. 26-28 May, 2010, Washington, DC, United States.
- Goodhart, C., & Hofmann, B. (2008). House prices, money, credit, and the macroeconomy. *Oxford Review of Economic Policy*, 24(1), 180–205.
- Greenwood, J., & Hercowitz, Z. (1991). The allocation of capital and the time over the business cycle. *Journal of Political Economy*, 99, 1188–1214.
- Grum, B., & Govekar, D. K. (2016). Influence of macroeconomic factors on prices of real estate in various cultural environments: case of Slovenia, Greece, France, Poland and Norway. *Procedia Economics and Finance*, 39, 597–604.
- Güner, Z. (2014). Cart and logistic regression analysis in data mining: an application on pharmacy provision system data. *Sosyal Güvence*, 6, 53–99.
- Hancke, G. P., & Malan, R. (1998). A modal analysis technique for the on-line particle size measurement of pneumatically conveyed pulverized coal. *IEEE Transactions on Instrumentation and Measurement*, 47(1), 114–122.
- Harris, R., & Arku, G. (2006). Housing and economic development: the evolution of an idea since 1945. *Habitat International*, 30(4), 1007–1017.
- Jayasekare, A. S., Herath, S., Wickramasuriya, R., & Perez, P. (2019). The price of a view: estimating the impact of view on house prices. *Pacific Rim Property Research Journal*, 25(2), 141–158.
- Kavzoğlu, T., & Çölkesen, İ. (2010). Investigation of the effects of kernel functions in satellite image classification using support vector machines. *Gebze High Technology Institute The Journal of Map*, 144, 73–82.
- Khalafallah, A. (2008). Neural network based model for predicting housing market performance. *Tsinghua Science and Technology*, 13(S1), 325–328.
- Kömürlü, R., & Önel, H. (2007). Source obtaining model approaches for house production in Turkey. *Megaron / Yıldız Technical University Faculty of Architecture E-Journal*, 2(2), 89–107.
- Lacoviello, M. (2000). House prices and the macroeconomy in Europe: results from a structural VAR analysis. ECB Working Paper 18, Frankfurt/Main.
- Leung, C. (2004). Macroeconomics and housing: a review of the literature. *Journal of Housing Economics*, 13(4), 249–267.
- Li, D. Y., Xu, W., Zhao, H., & Chen, R. Q. (2009). A SVR based forecasting approach for real estate price prediction. Paper presented at the 2009 International Conference on Machine Learning and Cybernetics, 12-15 July, 2009, Baoding, China.
- Li, L., & Chu, K.-H. (2017). Prediction of real estate price variation based on economic parameters. Paper presented at the 2017 International Conference on Applied System Innovation (ICASI), 13-17 May, 2017, Sapporo, Japan.
- Limsombunchai, V., & Samarasinghe, S. (2004). House price prediction using artificial neural network: a comparative study with hedonic price model. *Applied Economics Journal*, 9(2), 65–74.
- Mankiw, N. G., & Weil, D. N. (1989). The baby boom, the baby bust, and the housing market. *Regional Science and Urban Economics*, 19(2), 235–258.
- Mohd Radzi, M. S., Muthuveerappan, C., Kamarudin, N., & Mohammad, I. S. (2012). Forecasting house price index using artificial neural network. *International Journal of Real Estate Studies*, 7(1), 43–48.
- Nghiep, N., & Al, C. (2001). Predicting housing value: a comparison of multiple regression analysis and artificial neural networks. *Journal of Real Estate Research*, 22(3), 313–336.
- Osuna, E. E. (1998). Support vector machines: training and applications. Ph.D. dissertation, MIT, Department of Electrical Engineering and Computer Science, Cambridge, MA.
- Özkan, G., Yalpır, Ş., & Uygunol, O. (2007). An Investigation on the Price Estimation of Residable Real Estates by Using ANN and Regression Methods. The 12th Applied Stochastic Models and Data Analysis International Conference (ASMDA). Chania, Crete, Greece. 29 May-1 June 2007,
- Öztürk, N. & Fitöz, E. (2009). The determinants of housing sector in Turkey: an empirical analysis. *International Journal of Management Economics and Business*, 5(10), 21–46 .
- Pai, P. F., & Wang, W. C. (2020). Using Machine Learning models and actual transaction data for predicting real estate prices. *Applied Sciences*, 10(17), 5832.
- Painter, G., & Redfearn, C. L. (2002). The role of interest rates in influencing long-run homeownership rates. *The Journal of Real Estate Finance and Economics*, 25(2), 243–267.
- Poterba, J. M., Weil, D. N., & Shiller, R. (1991). House price dynamics: the role of tax policy and demography. *Brookings Papers on Economic Activity*, 1991(2), 143–203.

- Priemus, H. (2010). The credit crunch: Impacts on the housing market and policy responses in the Netherlands. *Journal of Housing and the Built Environment*, 25(1), 95–116.
- Rapach, D. E., & Strauss, J. K. (2007). Forecasting real housing price growth in the eighth district states. *Federal Reserve Bank of St. Louis: Regional Economic Development*, 3(2), 33–42.
- Selim, H. (2009). Determinants of house prices in Turkey: hedonic regression versus Artificial Neural network. *Expert Systems with Applications*, 36(2), 2843–2852.
- Takma, Ç., Gevrekçi, Y., Karahan, A. E., Atıl, H., & Çevik, M. (2017). Determining some traits effect on egg production of layers using Regression Tree Analysis. *Journal of Agriculture Faculty of Ege University*, 54(4), 459–463.
- Temür, A. S., Akgün, M., & Temür, G. (2019). Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models. *Journal of Business Economics and Management*, 20(5), 920–938.
- Vapnik, V. (1999). *The nature of statistical learning theory*. New York, Springer Science & Business Media.
- Wong, T. Y. J., Hui, C. M. E., & Seabrooke, W. (2003). The impact of interest rates upon housing prices: an empirical study of Hong Kong's market. *Property Management*, 21(2), 153–170
- Yılmazel, Ö., Afşar, A., & Çiçek, A. (2018). Using artificial neural network method to predict housing prices. *International Journal of Economics and Administrative Studies*, 20, 285–300.
- Zainun, N., Rahman, A. ve Eftekhari, M. (2010). Forecasting low-cost housing demand in Johor Bahru, Malaysia using Artificial Neural Network (ANN). *Journal of Mathematics Research*, 2(1), 14–19.