Advanced Geomatics

http://publish.mersin.edu.tr/index.php/geomatics/index

e-ISSN: 2791-8637

Housing Valuation Model in Samsun, Atakum District with Artificial Neural Networks and Multiple Regression Analysis

Mehmet Emin Tabar*[1], Aslan Cihat Başara [2], Yasemin Şişman [3]

[1]*Bitlis Eren University, Vocational School of Technical Sciences, Department of Architecture and Urban Planning, Bitlis, Turkey*
[2]*Ondokuz Mayis University, Institute of Graduate Studies, Department of Geomatics Engineering, Samsun, Turkey*
[3]*Ondokuz Mayis University, Faculty of Engineering, Department of Geomatics Engineering, Samsun, Turkey*

Keywords
Real estate appraisal
Multiple regression analysis
Artificial neural network
Artificial intelligence
Matlab

ABSTRACT

Valuation, in its simplest form, is the determination of the amount that a property will be processed at a certain date. Valuation can be done for many purposes. These; can be listed as buying and selling, transfer, tax assessment, expropriation, inheritance distribution, investment, financing and credit. There are various methods of valuation. These methods are examined under 3 main groups as traditional, statistical and modern valuation methods. The aim of the article is to provide an overview of regression analysis, one of the statistical valuation methods, and artificial neural networks, one of the modern valuation methods, and to compare the accuracy values. Matlab software was used for artificial neural network modeling and Minitab software was used for regression analysis. The accuracies of the obtained values were determined by the average absolute percent error (MAPE) formula.

1. INTRODUCTION

Real estate is independent and permanent rights registered on a separate page in the land register and independent parts of the property ownership register. Also, real estate gives owners the right to use as she wishes, except for the restrictions developed for the benefit of the public. (Açlar and Çağdaş 2002). Real estate valuation and these values reflection as tax are the most important economic foundations of developed societies. The real estate market has gained a positive momentum with the development of real estate investment trusts, construction companies, education, technology and professionalism. These developments have also brought professional real estate appraisals to the agenda in order to make the right investments (Atik et al. 2015). The house is the place where the consumer lives with families. (Özdamar 2004). The valuation process can be defined as determine the value of something measured with money. In this process, the attributes of something are valuated (Yomralıoğlu et al. 2011). Real estate valuation is made for trading or corporate transactions that varies according to needs, wishes and financial capacity. (Ring and Dasso 1977). According to another definition, it is the process of determining the provision of the seller according to the properties of the property for investment or long-term use (Brown 1965). Real estate valuation is done in many different ways. However, for professional real estate valuation, a mathematical model should be mentioned rather than subjective value estimates (Tabar and Şişman 2020). In traditional valuation methods, valuation experts try to make value calculations by only estimating an exchange price. Thus, the traditional methods are far from being objective compared to statistical and modern valuation methods since they do not depend on a mathematical model. Statistical and modern methods involve less initiative as they depend on a mathematical model. The most important issues in the valuation area is the need to that the information provided to the customer is clear. (Pagourtzi et al. 2003). When the methods used in the mathematical model are examined; many methods such as fuzzy logic, artificial neural networks, spatial analysis, support vector machines, regression analysis are reached. The most used real estate valuation methods are shown in Table 1.

* **Corresponding Author**

*(metabar@beu.edu.tr) ORCID ID 0000 – 0002 – 3234 – 5340
(aslancihatbasara@gmail.com) ORCID ID 0000 – 0001 – 6644 – 6097
(ysisman@omu.edu.tr) ORCID ID 0000 – 0002 – 6600 – 0623

Cite this article;

Tabar M E, Basara A C & Sisman Y (2021). Housing Valuation Model in Samsun, Atakum District with Artificial Neural Networks and Multiple Regression Analysis. Advanced Geomatics, 1(1), 27-32

Research Article

Received: 04/09/2021; Accepted: 30/09/2021

Table 1. The most used real estate valuation methods

Traditional methods	Statistical methods	Modern methods
Comparison method	Nominal method	Fuzzy logic
Income method	Multiple regression method	Artificial neural networks
Cost method	Hedonic pricing method	Support vector machine

In this study, multiple regression analysis, one of the statistical valuation methods, and artificial neural networks, one of the modern valuation methods, were used. Accuracy of methods were calculated with the formula of average absolute percent error (MAPE). The comparison was made according to the MAPE of methods.

2. MATERIAL AND METHOD

2.1. Material

Atakum district is one of the most important and most preferred districts in Samsun in terms of real estate valuation. Atakum is listed as the 23rd district in the most housing sales ranking among 923 districts of Turkey in 2019. A total of 10607 housing properties were sold in Atakum in 2019. This study in Samsun, Yenimahalle Neighborhood, a real estate valuation model was created using artificial neural networks and multiple regression analysis and the value was estimated. The study area is shown in Fig 1.

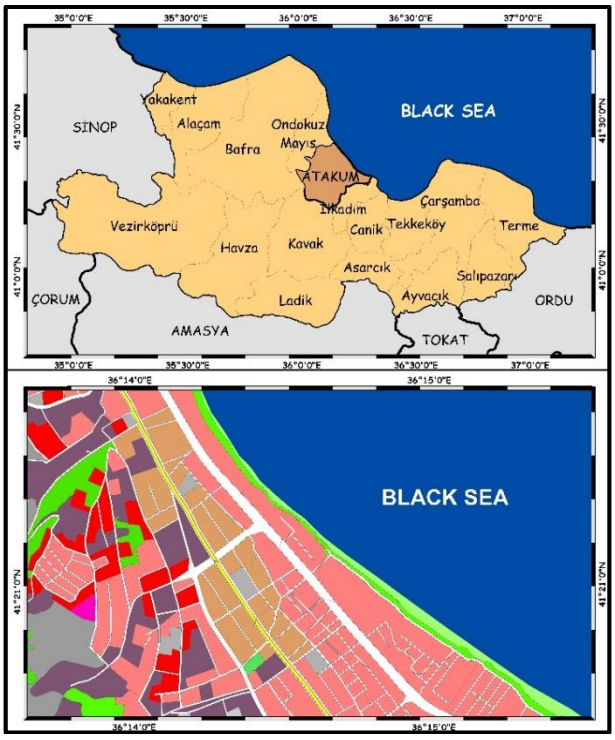

Figure 1. Study area

The data were obtained from a housing sales site. While choosing the housing, care was taken to choose those with a facade to the tramway street. Housing prices are valid from August to October 2020. Housing prices were determined by modeling the obtained data with multiple regression analysis and artificial neural networks. Minitab software was used for this process. The estimated values after modeling were compared with the housing values.

2.2. Method

2.2.1. Multiple Regression Analysis

Multiple regression analysis consists of response and independent variables. The number of response variables is single, but the number of arguments can be more than one. The relationship between response and independent variables can be defined as linear curvilinear exponential etc., If there is only one independent variable and the relationship is linear, it is called simple linear regression, if there are two or more independent variables and the relationship is linear, it is called multiple linear regression. In regression analysis, it is aimed to make the relationship between variables functionally meaningful and explain this relationship with a model (Chatterjee and Hadi 2015).

In valuation models, more than one variable can be combined to form a variable. These variables can also affect each other. Therefore, single regression analysis is not possible. More than one analysis using variables is called multiple regression analysis (Karacabey and Gökgöz 2012). In multiple regression, there are more than one independent variable affecting the response variable. These types of studies have two general purposes:

- Which of the independent variables or finding out which ones affect the response variable more.
- Response variable with the help of independent variables estimate.

The mathematical model of the regression method is as follows:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots\ldots\ldots + \beta_n X_{ni} + \varepsilon$$

Yi: response variable
Xi: independent variable
i = 1, 2, ..., n

β_0: Returns the value of the point where the line intersects the y-axis. It takes names like intercept.

β_i: It is the slope of the line. It takes names such as slope, speed, regression coefficient (Durmuş 2016).

The result of the regression analysis is tested with some values as determination and correlation coefficient. The determination coefficient is named as the R^2 (Sisman 2014). The R^2 indicates the goodness of fit for the model (Sisman et al. 2014). The R^2 value is a measure of linearity. The range of R^2 is between 0 and +1. If R^2 value is close to +1, the observation values means it's dispersing around the line (Efe et al. 2000). The correlation coefficient is indicating the relationship between two statistically. The range of correlation coefficient is between -1 and +1. If the correlation coefficient is" 0 ", there is no relationship between X and Y. Also, If the correlation coefficient is close to " +1 or -1 " there is positive or negative relationship between X and Y (Güngör and Sevindir 2013).

2.2.2. Artificial Neural Networks

Artificial neural networks emerged by artificially imitating the way the human brain works. It can be thought of as a complex system that occurs as a result of connecting many nerve cells in the brain with different levels of influence. In artificial neural networks, the system first performs the learning process by analyzing input data and output data (Öztürk and Şahin 2018). It gives approximate outputs of new input data after learning process as a result of iterations. Artificial neural networks are especially used in engineering applications. Engineering problems that are difficult to solve with classical methods have gained a different dimension with artificial neural networks and have created an effective alternative (Yegnanarayana 2009). Although the human brain is limited in mathematical operations such as division, multiplication, addition, subtraction, it is more successful than machines in many processes such as learning, remembering, and predicting. The main features of artificial neural networks are nonlinearity, learning, parallel working, generalization, working with missing data, using a large number of variables and parameters, applicability, fault tolerance and flexibility.

Artificial neural networks consist of 3 main components. These; architectural structure, learning algorithm and activation function. When we examine the architectural structure, the input layer consists of the hidden layer and the output layer. In the learning algorithm, the weights in the whole network should take optimal values. In fact, training the net is to find the best value of the weights (Graupe 2013). The activation function provides the match between input and output layers.

Artificial neural networks learn by making mistakes. Basically, artificial neural networks learn in 3 stages. In the first step, outputs are calculated. In the second step, it compares the outputs with the target outputs and calculates the error. In the last stage, the process repeats by changing the weights (Livingstone 2008). Architecture of multilayer artificial neural network is shown in Fig 2 as input layer, hidden layer and output layer.

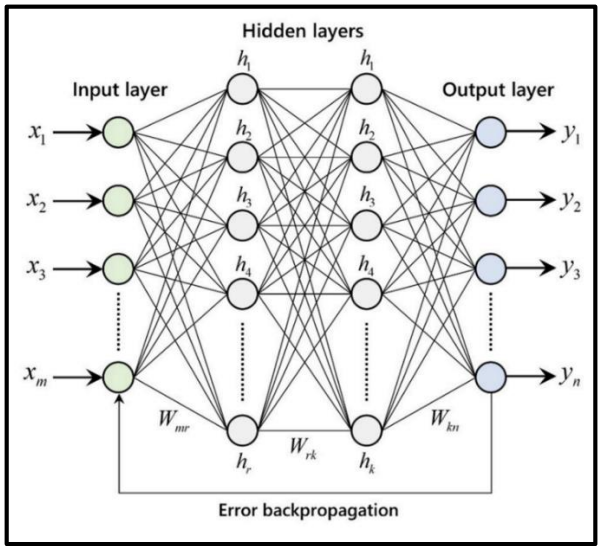

Figure 2. Architecture of multilayer artificial neural network (Fernández-Cabán et al. 2018)

Artificial neural networks consist of multi-layer computational units. The data received from the external environment is applied to the input layer and processed at the entrance, the information is transmitted to the middle layer without any change in the flow direction. The information entering the process in these layers is transmitted forward to the output layer. Artificial neural networks based on the working principle of transmitting information in the forward direction from input to output are called forward feed artificial neural networks (Canan 2006).

3. APPLICATION AND RESULTS

In Samsun, Atakum, Yenimahalle Neighborhood, 200 data were collected about the properties and values of the houses facing the tramway. Housing data are generally shown in Table 2.

Table 2. Housing data

Area	Number of rooms	Building age (Years)	Floor/Number of floors	Number of bathrooms	Balcony	Furnished	Value (TRY)
45	1+1	0	0/4	1	Yes	Yes	205 000
125	3+1	21-25	0/4	2	Yes	No	300 000
160	4+1	11-15	8/8	2	Yes	No	375 000
85	2+1	0	-1/4	1	Yes	No	260 000
130	3+1	5-10	5/8	2	Yes	No	435 000
…	…	…	…	…	…	…	…
130	3+1	4	1/5	1	Yes	No	349 000
145	3+1	0	1/6	2	Yes	No	475 000
135	3+1	0	1/6	2	Yes	No	495 000
145	3+1	16-20	6/6	1	Yes	No	290 000
95	2+1	11-15	0/7	2	Yes	Yes	260 000

These values were transformed into tables and normalized with maximum minimum normalization.

$$Normalized\ data = \frac{(x - \min)}{max - min}$$

For the multiple regression analysis, the normalized data were defined in Minitab program. The result of the analysis was taken as a mathematical model and compared with the actual values of the test data. For the artificial neural network analysis, the artificial neural network module in Matlab software was used. The normalized housing data were defined as input, output and test data. A feed forward network was created using this module. For the test data, the outputs of the Matlab software are taken and compared with the real housing values.

A comparison was made by calculating the accuracy of models using MAPE.

$$MAPE = 100\frac{\sum_{i=1}^{n}\frac{|A_i - F_i|}{A_i}}{n}$$

In this formula, A_i is the real value and F_i is the predicted value.

3.1. Multiple Regression Analysis Application

Normalized housing variables were defined in Minitab software and regression equation was obtained. Housing values were obtained as follows;

Housing Value = (-0.612 + 0.405 area + 0.307 room + 0.4227 b.age + 0.1027 floor + 0.1712 bathroom + 0.2276 balcony - 0.0229 furnished)

The calculated normalized value has been converted into the real house value with the maximum-minimum formula.

Normal probability plot created according to the value of the house is shown in Fig 3.

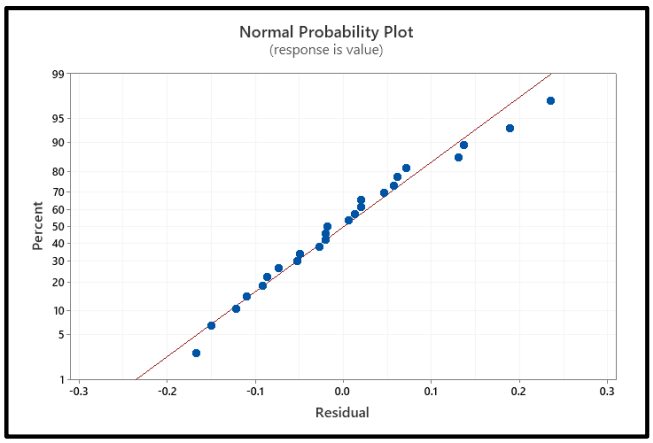

Figure 3. Normal probability plot

The values calculated by multiple regression analysis are shown in Table 3 together with the housing sale prices.

Table 3. Multiple regression analysis results

MRA Value (TRY)	Sale Value (TRY)	Accuracy (%)
219361.667	205000.000	92.994
294014.667	300000.000	98.004
410175.000	375000.000	90.620
286537.000	260000.000	89.793
380271.000	435000.000	87.418
...	...	...
331190.000	349000.000	94.896
437010.667	475000.000	92.002
426800.667	495000.000	86.222
295148.667	290000.000	98.224
267976.333	260000.000	96.932

3.2. Artificial Neural Networks Application

Normalized housing data are defined in Matlab software. The training of the network was carried out with 8 neurons using experimental data with the created feed forward artificial neural network. The training process was repeated a few times to make the learning process more accurate. The maximum failure value was entered as 500, and the iteration amount was determined as 1000.

The accuracy and consistency values of the data were examined by looking at the regression chart after the training. The trained network is simulated with test data.

Housing values were determined by applying the maximum-minimum normalization reversed to the values obtained. Artificial neural network training regression is shown in Fig 4. The results are shown in Table 4.

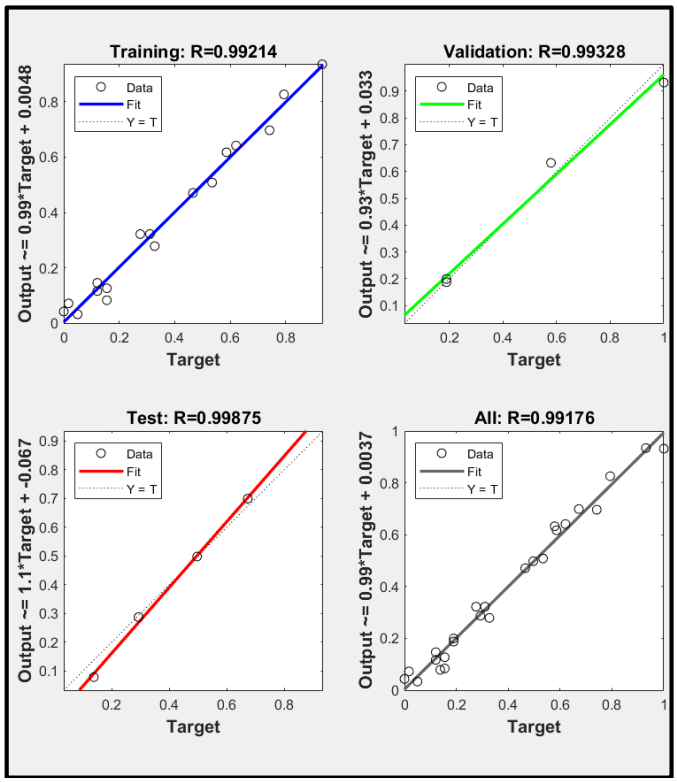

Figure 4. Artificial neural network training regression

Table 4. Artificial neural network results

ANN Value (TRY)	Sale Value (TRY)	Accuracy (%)
222862.080	205000.000	91.286
293771.863	300000.000	97.923
376021.579	375000.000	99.727
257108.193	260000.000	98.887
454902.955	435000.000	95.424
...	...	...
320026.201	349000.000	91.698
471619.035	475000.000	99.288
471124.399	495000.000	95.177
291098.107	290000.000	99.621
259925.188	260000.000	99.971

4. DISCUSSION

In the study, the values obtained in the housing valuation model created using multiple regression analysis and artificial neural networks were compared with the real values of the houses and the accuracy of methods were determined. The housing data was taken used in the study were obtained from Samsun province Atakum district Yenimahalle neighborhood. In determining the housing, the ones facing the tramway street were chosen as the location.

The total accuracy value obtained in the multiple regression analysis was determined as 93.232%. The total accuracy value obtained in artificial neural networks was determined as 97.090%.

5. CONCLUSION

Considering the accuracy values obtained in the study, it is seen that the housing valuation model created with artificial neural networks gives a higher accuracy value. Multiple regression analysis is more suitable in terms of ease of application and understandability of the model. In artificial neural networks, there may be some results that vary with the change of the number of neurons and the selection of functions. The authors recommend the use of statistics and modern methods in frequent applications such as real estate appraisal.

Author contributions

1st Author: Conceptualization, Methodology, Software, Data Curation, Writing-Original Draft Preparation, Validation, Visualization
2nd Author: Visualization, Data Curation
3rd Author: Investigation, Reviewing and Editing

Conflicts of interest

The authors declare no conflicts of interest.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

Açlar A & Çağdaş V (2002). Taşınmaz (gayrimenkul) değerlemesi. TMMOB Harita ve Kadastro Mühendisleri Odası, Ankara.

Atik M, Köse Y, Yilmaz B & Erbaş M (2015). Şehirlerin İlerleme Yönlerinin Gayrimenkul Değerleri Üzerindeki Etkisinin Ölçülmesi. Çankırı Karatekin Üniversitesi, 5, 443-458.

Brown R K (1965). Real estate economics: an introduction to urban land use: Houghton Mifflin.

Canan S (2006). Yapay sinir ağları ile GPS destekli navigasyon sistemi. Selçuk Üniversitesi Fen Bilimleri Enstitüsü.

Chatterjee S & Hadi A S (2015). Regression analysis by example: John Wiley & Sons.

Durmuş B (2016). Konut Fiyatlarını Etkileyen Parametrelerin Çoklu Regresyon Analizi Yöntemiyle İrdelenmesi Ve Kentsel Dönüşüme Katkıları. Fen Bilimleri Enstitüsü.

Efe E, Bek Y & Şahin M (2000). SPSS'te Çözümleri ile İstatistik Yöntemler II, Kahramanmaraş Sütçü İmam Üniversitesi Rektörlüğü Yayın No: 73, Ders Kitapları Yayın No: 9, KS Ü. Basımevi, Kahramanmaraş, 214.

Fernández-Cabán P L, Masters F J & Phillips B M (2018). Predicting roof pressures on a low-rise structure from freestream turbulence using artificial neural networks. Frontiers in Built Environment, 4, 68.

Graupe D (2013). Principles of artificial neural networks (Vol. 7): World Scientific.

Güngör A & Sevindir H C (2013). Isparta İlindeki Atmosferde Bulunan Kükürt dioksit (SO2) ve Partikül Madde (PM) Konsantrasyonunun Çoklu Doğrusal Regresyon Yöntemi İle Modellenmesi. Journal of Natural and Applied Science, 17(1), 95-108.

Karacabey A & Gökgöz F (2012). Çoklu Regresyon Modeli. Anova Tablosu, Matrislerle Regresyon Çözümlemesi, Regresyon Katsayılarının Yorumu.

Livingstone D J (2008). Artificial neural networks: methods and applications: Springer.

Özdamar N (2004). 4822 Sayılı Yasa İle Değişik 4077 Sayılı Yasa'da Tanımlanan Konut Nedir. Türkiye Barolar Birliği Dergisi, Yıl, 17, 317-331.

Öztürk K & Şahin M E (2018). Yapay Sinir Ağları ve Yapay Zekâ'ya Genel Bir Bakış. Takvim-i Vekayî, 6(2), 25-36.

Pagourtzi E, Assimakopoulos V, Hatzichristos T & French, N (2003). Real estate appraisal: a review of valuation methods. Journal of Property Investment & Finance.

Ring A A & Dasso J J (1977). Real estate principles and practices: Prentice Hall.

Sisman Y (2014). The optimization of GPS positioning using response surface methodology. Arabian Journal of Geosciences, 7(3), 1223-1231.

Sisman Y, Elevli S & Sisman A (2014). A statistical analysis of GPS positioning using experimental design. Acta Geodaetica et Geophysica, 49(3), 343-355.

Tabar M E & Şişman Y (2020). Bulanık Mantık ile Arsa Değerleme Modelinin Oluşturulması. Türkiye Arazi Yönetimi Dergisi, 2(1), 18-24.

Yegnanarayana B (2009). Artificial neural networks: PHI Learning Pvt. Ltd.

Yomralıoğlu T, Nişancı R Çete, M & Candaş E (2011). Dünya'da ve Türkiye'de Taşınmaz Değerlemesi.