*Journal of Statistical and Econometric Methods, Vol. 9, No. 4, 2020, 31-38*
*ISSN: 2241-0384 (print), 2241-0376 (online)*
*Scientific Press International Limited*

# Predicting House Prices in Turkey by Using Machine Learning Algorithms

**MEHMET Erkek[1], KAMİL Çayırlı[2] and ALİ Hepşen[3]**

## Abstract

Housing prices are an important reflection of the economy, and housing price ranges are of great interest for both buyers and sellers. The goal of this paper is to empirically conduct the best machine learning regression model for Turkish Housing Market by comparing accuracy scores and absolute deviations of test results by using Python programming language and Keras library for the five-year period from January 2015 to December 2019. This study consists of 15 explanatory variables describing (almost) every aspect of houses in Istanbul, Izmir and Ankara. These fifteen explanatory building and dwelling variables are used for each prediction model. In this study, three different data models are created by using support vector machine, feedforward neural networks and generalized regression neural networks algorithms. The experiments demonstrate that the Feedforward Neural Network model, based on accuracy, consistently outperforms the other models in the performance of housing price prediction. According to another result of the study, the most important variables in the model are the location of the house and the size of the house, while the size of the terrace is determined as the least important variable.

**JEL classification numbers:** R10, R15, R19
**Keywords:** Housing market, Zingat.com, Machine learning, House price prediction, Python programming language, Keras library

[1] Zingat.com.
[2]. Vocational School, Nisantasi University.
[3] Faculty of Business Administration, Istanbul University.

Article Info: *Received:* March 5, 2020. *Revised*: May 1, 2020.
*Published online*: August 30, 2020.

## 1. Introduction

Housing is accepted as a fundamental right in the international arena according to the Universal Declaration of Human Rights (1948). Moreover, HABITAT II Conferences organized by United Nations have underlined the importance of housing with a main theme of UN Habitat Agenda as "enough housing for everyone" (Habitat II, 1996) and housing matters as a major quality of life issue. On the other hand, it is a relatively illiquid investment, with an uncertain capital value, and it is generally highly leveraged, which makes it a potentially important channel of transmission of monetary policy.

Especially in Turkey, due to population growth, decrease in average size of the household, migration and rapid pace of urbanization, housing demand is continuously increasing, and since this requirement cannot be met with supply, house prices have been increasing steadily since 2000s. But when looking towards to future, to predict the house prices and to monitor the evolution of the market trend over time, is obviously a necessary requirement in Turkey. With the help of artificial intelligence, the goal of this paper is to empirically conduct the best machine learning regression model for Turkish Housing Market by comparing accuracy scores and absolute deviations of test results by using Python programming language and Keras library.

The layout of the remainder of this article is as follows. The next section presents a literature review between machine learning and the real estate. The following section provides overview of support vector machines, feedforward artificial neural networks, generalized regression neural networks, and the data. Section four describes the data analysis and presentation of findings. The final section is the conclusion.

## 2. Literature Review

The relationship between machine learning and the real estate market has come to prominence in recent years, especially with the development of technology and studies on data analysis. Some previous studies have been made to analyze the area of pricing of real estate, with neural networks or other forms of machine learning. There was an article on the appraisal of real estate values in Lagos, Nigeria. An artificial neural network (ANN) was applied in property valuation using the Lagos metropolis property market as a representative case (Abidoye and Chan, 2017). Property sales transactions data (11 property attributes and property value) were collected from registered real estate firms operating in Lagos. The result showed that the ANN model possesses a good predictive ability, implying that it was suitable and reliable for property valuation. The relative importance analysis conducted on the property attributes revealed that the number of servants' quarters is the most important attribute affecting property values. The findings suggested that the ANN model could be used as a tool by real estate stakeholders, especially appraisers and researchers for property valuation.

Another study conducted a study on the use of neural networks to predict the prices of housing in Singapore (Wang et al., 2016). This network uses time-series data of housing sales, as well as general economic variables, such as the population levels and the average monthly wages. The authors praised the results, claiming the model delivers an accurate forecast. This was based on a high measure of the coefficient of determination (r2), and a low mean square error of the prediction, which suggests a good performing model. However, the authors did not provide any mean percentage errors, making comparisons to the other mentioned studies difficult. Machine learning techniques were applied to analyze the purchase and sale of historic properties in Australia to find useful models for home buyers and sellers (Phan, 2018). The result of the study was the high inconsistency between the most expensive and affordable home prices in Melbourne. Furthermore, experiments have shown that the combination of Stepwise and Support Vector Machines based on the mean square error measurement is a competitive approach.

A prognostic model of real estate market value in EU countries established due to the impact of macroeconomic indicators (Ćetković et al., 2018). The available inputs have shown that macroeconomic variables affect the determination of property prices. In another study, various regression techniques were used and the results were calculated as the weighted average of various techniques (Varma et al., 2018). The results showed that the approach provides minimum error and maximum accuracy compared to the individual algorithms applied. Authors of the study proposed the Genetic Algorithm, Geographical Information, Support Vector Machine model for the estimation of housing prices and combines artificial neural networks with Particle Swarm Optimization (Zhou et al., 2018). In this study, they focused on the current literature on artificial neural networks. Finally, proposals were made for the assessment of housing in China.

New Support Vector Machine (SVM) approach was applied for the estimation of house prices (Chen et al., 2017). According to them, the results calculated with non-parametric estimates confirm that the predictive power of using SVM in the house price estimation is highly accurate. In addition, studies to extract geographic weights using core density estimates to reflect price responses to quantities of local hedonic properties are recommended. Artificial Neural Network (ANN) was used in real estate appraisal using the Lagos metropolitan real estate market as a representative case (Abidoye and Chan, 2017). The results showed that the ANN model has a good estimation ability, and it was suitable and reliable for property valuation and could be used as a tool by the housing market, especially valuation and researchers for pricing. A recent study tried to investigate the potential of using neural networks to predict selling prices of apartments in Stockholm, based on apartment parameters (Nilsson, 2019). In his study, networks were trained to either make an improved valuation, based on a listing price, or make a new valuation of an apartment. The results were promising, and in line with contemporary findings; however, the worst-case performance of the models could make them unsuitable for many purposes.

## 3. Theoretical Framework and Data

This section presents the details concerning the data and algorithms used in the study. The choice of the algorithms to perform housing prices prediction has taken into account the nature of the dataset, since real-world data are not always balanced, complete, scaled, or easy to handle (Alpaydin, 2020). Support vector machines, feedforward artificial neural networks and generalized regression neural networks are used in the development of house price prediction models.

Performance of the support vector machines and the developed models for house price estimation are directly proportional to the determination of the optimal values of the kernel function and the parameters of the function. In this study, Radial, Linear, Polynomial and Sigmoid functions were tested as kernel functions. The parameter values of the specified kernel functions and the optimum values of the Epsilon and cost parameters of the Support Vector Machines were determined by grid search.

The performance of house price estimation models developed with Feedforward Neural Networks depends on many parameters such as number of hidden layers to be used in the network; number of neurons to be found in hidden layers; number of epokes; learning coefficient; momentum and activation functions to be used in hidden/ exit layers. While estimation models were formed, optimum values of these parameters were found during the training phase. The training process on the dataset continued until a certain number of epokes were reached or below a predetermined threshold error value.

The performance of real estate valuation estimation models to be developed with Generalized Regression Neural Networks is used as kernel function. It depends on the parameter value of the Gaussian function. The optimum value of this parameter was determined during the training phase of the data set and neuron pruning was also applied to improve the performance of this method.

Due to the large number of estimation variables affecting the property price, whether more accurate estimates can be obtained with fewer estimation variables by using Relief-F and Minimum Redundancy Maximum Relevance (mRMR) methods was investigated. Mean Absolute Percentage Error (MAPE), which is frequently used in the literature, was used to evaluate the performance of estimation models. Python programming language and Keras library were also used to construct prediction models.

In this paper we used the locational listing data of Turkish leading property portal Zingat.com[*], for the five-year period from January 2015 to December 2019. With 15 explanatory variables (location (latitude& longitude), dwelling size, terrace size, number of room, number of bathroom, building age, total floor, floor number, property type, view, view direction, lift, heating system, availability of security, and parking) describing (almost) every aspect of residential homes in Istanbul, Izmir

[*] Zingat.com is meeting point and source of information for investors in the real estate sector, real estate sellers / buyers, people who is looking for a new house, professionals who earn in this sector. In this respect, it aims to provide accurate real estate appraisal services to its users.

and Ankara that are the biggest cities in Turkey. These fifteen explanatory building and dwelling variables were used for each prediction model. Preprocessing operations (clearing of outliers, missing data completion, etc.) were also performed on the data sets and the data set was divided into two as sale listings and rental listings. The following tables (Table 1 and Table 2) contain detailed data sets:

**Table 1: Housing Prices for Sale**
**(Average Price for 100Sqm House for the years 2015-2019)**

| id | City | County | District | Status | Price (TL) |
|---|---|---|---|---|---|
| 1 | İstanbul | Beylikdüzü | Merkez | Sales | 320.000 |
| 2 | Ankara | Çankaya | İlkbahar | Sales | 365.696 |
| 3 | Ankara | Çankaya | Çankaya | Sales | 256.292 |
| 4 | İstanbul | Beylikdüzü | Dereağzı | Sales | 293.893 |
| 5 | İstanbul | Beylikdüzü | Marmara | Sales | 315.200 |
| 6 | İstanbul | Beylikdüzü | Yakuplu | Sales | 400.000 |
| 7 | İstanbul | Beylikdüzü | Marmara | Sales | 280.000 |
| 8 | İstanbul | Beylikdüzü | Adnan Kahveci | Sales | 300.653 |
| 9 | İstanbul | Beylikdüzü | Büyükşehir | Sales | 272.000 |
| 10 | İstanbul | Beylikdüzü | Kavaklı | Sales | 345.000 |
| 11 | İstanbul | Beylikdüzü | Gürpınar | Sales | 300.653 |
| 12 | İstanbul | Beylikdüzü | Cumhuriyet | Sales | 171.523 |
| 13 | Ankara | Çankaya | Oran | Sales | 490.668 |
| 14 | İzmir | Beylikdüzü | Kavaklı | Sales | 265.955 |
| 15 | Ankara | Çankaya | Ayrancı | Sales | 360.540 |

**Table 2: Housing Prices for Rent (Average Prices for the years 2015-2019)**

| id | City | County | District | Status | Price (TL/SQM) |
|---|---|---|---|---|---|
| 1 | Ankara | Çankaya | Oran | Rental | 16,50 |
| 2 | Ankara | Çankaya | Büyükesat | Rental | 13,45 |
| 3 | Ankara | Çankaya | Ayrancı | Rental | 9,00 |
| 4 | Ankara | Çankaya | Ayrancı | Rental | 7,99 |
| 5 | İstanbul | Beylikdüzü | Dereağzı | Rental | 6,00 |
| 6 | Ankara | Çankaya | Sancak | Rental | 7,00 |
| 7 | İzmir | Karşıyaka | Mavişehir | Rental | 9,00 |
| 8 | Ankara | Çankaya | Birlik | Rental | 6,60 |
| 9 | Ankara | Çankaya | Gaziosmanpaşa | Rental | 6,50 |
| 10 | Ankara | Çankaya | Çankaya | Rental | 5,50 |
| 11 | İzmir | Karşıyaka | Yalı | Rental | 2,50 |
| 12 | Ankara | Çankaya | Çankaya | Rental | 5,50 |
| 13 | İstanbul | Beylikdüzü | Barış | Rental | 6,50 |
| 14 | Ankara | Çankaya | Oran | Rental | 6,00 |
| 15 | İzmir | Karşıyaka | Mavişehir | Rental | 8,00 |

# 4. Data Analysis and Presentation of Findings

## 4.1 House Prices Forecast Model Selection

The data was split into training, and testing sets. The 80-20 split used is a typical ratio for this purpose; 80% of the data were considered as training set and 20% as test set. To evaluate the performance three different models have been formed with the support vector machines, feedforward artificial neural networks and generalized regression neural networks. Phyton programming language and Keras library were used in the analysis.

Mean Absolute Percentage Error (MAPE) was selected as the error metric. MAPE is a measure of prediction accuracy of a forecasting method in statistics, for example in trend estimation, also used as a loss function for regression problems in machine learning. It usually expresses the accuracy as a ratio defined by the formula:

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_{t-}F_t}{A_t}\right| \tag{1}$$

where $A_t$ is the actual value and $F_t$ is the forecast value. The MAPE is also sometimes reported as a percentage, which is the above equation multiplied by 100. The difference between $A_t$ and $F_t$ is divided by the actual value At again. The absolute value in this calculation is summed for every forecasted point in time and divided by the number of fitted points n. Multiplying by 100% makes it a percentage error.

When the success of the methods is examined according to the results of the analysis, it has been found that the artificial neural network model gives better results than the other supervised learning models.

**Table 3. Model Comparison**

| MODEL | GRNN MAPE | MLP MAPE | SVM MAPE |
|---|---|---|---|
| Rental | 12,1% | 9,3% | 11,49% |
| Sales | 19,1% | 14,9% | 17,93% |

### 4.2 Feedforward Artificial Neural Network

A Multilayer Perceptron (MLP) Network consisting of a total of 34 cells consisting of 1 input, 2 hidden and 1 output layers, 15 in the input layer, 9 in the each hidden layers and 1 in the output layer was formed. A rectified linear unit (ReLU) was used in the middle layers of the network and linear function was used in the output layer. Epoch number = 1000, Learning Rate = 0.01, Momentum Rate = 0.9 and Optimization function = Adam were used. MAPE in the study was 9.3% in rental listings; 14.9% for sale listings.

## 5. Conclusion

In this study, we tried to conduct the best machine learning regression model for Turkish Housing Market by comparing accuracy scores and absolute deviations of test results. In our study, fifteen explanatory building and dwelling variables were used for each prediction model; Location (latitude& longitude), Dwelling Size, Terrace Size, Number of Room, Number of Bathroom, Building Age, Total Floor, Floor Number, Property Type, View, View Direction, Lift, Heating System, Security, and Parking. Dataset was separated into two as rental listings and for sale listings.

The data was split into training, and testing sets. The 80-20 split used is a typical ratio for this purpose; 80% of the data were considered as training set and 20% as test set. To evaluate the performance three different models have been formed with the support vector machines, feedforward artificial neural networks and generalized regression neural networks. Python programming language and Keras library were used. Mean Absolute Percentage Error (MAPE) was selected as the error metric.

In our empirical part, when the 15 input variables of the model were examined, it was observed that the most important variable in the model was location. In the second place is the size of the apartment. The lowest weight belongs to number of balconies. All of these results show that artificial neural network model can be used as an important tool in estimating house prices.

## References

[1] P. Nilsson, "Prediction of Residential Real Estate Selling Prices Using Neural Networks", Master Thesis, KTH Royal Institute of Technology School of Electrical Engineering and Computer Science, 2019.

[2] T. D. Phan, "Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia", 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), Sydney, Australia, 2018, pp. 35-42.

[3] J. Ćetković et al., "Assessment of the Real Estate Market Value in the European Market by Artificial Neural Networks Application", Complexity, vol.4, no.1, 2018, pp. 1-10.

[4] A. Varma et al., "House Price Prediction Using Machine Learning and Neural Networks," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1936-1939.

[5] G. Zhou et al., "Artificial Neural Networks and the Mass Appraisal of Real Estate", International Journal of Online and Biomedical Engineering (iJOE), Vol.14, No.3, 2018, pp.180-187. https://online-journals.org/index.php/i-joe/article/view/ 8420

[6] J.H. Chen et al., "Forecasting Spatial Dynamics of the Housing Market Using Support Vector Machine", International Journal of Strategic Property Management, vol. 21, no. 3, 2018, pp. 273-283.

[7] United Nations, "HABITAT II Conference Conclusion Report", 1996.

[8] E. Alpaydin, "Introduction to Machine Learning", Fourth Edition, Adaptive Computation and Machine Learning Series, MIT Press, March 2020.

[9] R.B. Abidoye and A.P.C. Chan, "Modelling Property Values in Nigeria Using Artificial Neural Network", Journal of Property Research, vol. 34, no. 1, pp. 36-53, 2017.

[10] P. Wang et al., "Predicting Public Housing Prices Using Delayed Neural Networks," TENCON 2016 - 2016 IEEE Region 10 Conference, Singapore, 2016.