BŞEÜ Fen Bilimleri Dergisi
9(1), 138-151, 2022

BSEU Journal of Science
https://doi.org/10.35193/bseufbd.998331

e-ISSN: 2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)

Araştırma Makalesi - Research Article

Housing Price Estimation with Deep Learning: A Case Study of Sakarya Turkey

Derin Öğrenme ile Konut Fiyat Tahminlemesi: Türkiye'deki Sakarya İli Üzerinde Bir Çalışma

Murat Özdemir[1*], Kazım Yıldız[2], Büşra Büyüktanır[3]

Geliş / Received: 21/09/2022 ***Revize / Revised: 15/02/2022*** ***Kabul / Accepted: 17/02/2022***

ABSTRACT

Shelter is one of the most basic human needs. Besides housing needs, the housing market is also very important for investment. It is also a market where many people, such as engineers, architects, real estate agents make economic gain. When a house is bought for living in it, it is not desired to be changed for many years, and when it is bought for investment, it is a tool that requires good income. Therefore, the best decision should be made when buying a house, and it should be scrutinized. Correct estimation of house prices is very important for both buyers to make the right decision and for sellers to sell without a loss. There are many parameters for estimating house prices. In addition to variables such as the number of floors, location, and several bathrooms used in previous studies, economic factors (such as the price of bread, foreign currency price, new car price) and the housing loan interest rate of the banks were taken as inputs in this study. Sakarya province, where all parameters can be tested to make a more accurate determination, was chosen as the research area. A comparison of polynomial regression, random forest, and deep learning methods was made and it was concluded that the most accurate method was deep learning. At the same time, it was determined which parameters are more effective in house price estimation.

Keywords- Housing Price Prediction in Sakarya, Deep Learning, Random Forest, Polynomial Regression

ÖZ

Barınma insanların en temel ihtiyaçlarından biridir. Konut piyasası barınma ihtiyacını karşılamasının yanı sıra yatırım için de çok önemlidir. Aynı zamanda mühendis, mimar, emlakçı vb. çok sayıda insana da ekmek kapısı olmuştur. Konut, oturumluk için alındığında uzun yıllar değiştirilmek istenmeyen, yatırımlık için alındığında ise iyi kazanç elde edilmek istenen bir araçtır. Bu yüzden konut satın alınacağı zaman talebe göre doğru araştırma yapılmalı nihayetinde en doğru karar verilmelidir. Konut fiyatlarının doğru tahmin edilmesi hem alıcılar için doğru kararı verebilmek, hem de satıcılar için zarar etmeden satış yapabilmek için çok önemlidir. Konut fiyat tahmini için çok sayıda parametre vardır. Bunlardan daha önceki çalışmalarda da kullanılan kat sayısı, konum, banyo sayısı vb. değişkenlerin yanı sıra bu çalışmaya özgün olarak döneme ait ekonomik etkenler (ekmek fiyatı, döviz fiyatı, sıfır araç fiyatı gibi) ve o dönemde bankaların konut kredisi faiz oranı girdi olarak alınacaktır. Daha doğru tespit yapabilmek adına tüm parametrelerin test edilebileceği Sakarya ili araştırma alanı olarak seçilmiş ve yöntem olarak polinomal regresyon, rastgele orman ve derin öğrenme yöntemleri kullanılarak bu yöntemlerin kıyaslaması

[1]*Corresponding author contact: muraozdemir20@marun.edu.tr (https://orcid.org/0000-0001-7225-3574)
Computer Engineering, Institute of Science, Marmara University, Istanbul
[2]Contact: kazim.yildiz@marmara.edu.tr (https://orcid.org/0000-0001-6999-1410)
Computer Engineering, Faculty of Technology, Marmara University, Istanbul
[3]Contact: busra.buyuktanir@marmara.edu.tr (https://orcid.org/0000-0003-2571-4029)
Computer Engineering, Faculty of Technology, Marmara University, Istanbul

BŞEÜ Fen Bilimleri Dergisi
9(1), 138-151, 2022

BSEU Journal of Science
https://doi.org/10.35193/bseufbd.998331

e-ISSN:2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)

yapılarak en doğru yöntemin derin öğrenme olduğu sonucuna varılmış ve parametrelerin ev fiyatlarına etkileri tespit edilmiştir.

Anahtar Kelimeler- Sakarya Konut Fiyat Tahmini, Yapay Sinir Ağları, Rastgele Orman, Polinomal Regresyon

I. INTRODUCTION

Shelter is one of the most basic needs of people, as well as a very important investment tool. The housing market has an important place in the country's economy. Buying and selling a home is not a more frequent purchase like cars or appliances. People want to find the most suitable residence for them and live here for many years. The housing market is also an important market for employment. Many people, such as construction workers, architects, engineers are employed from the construction stage to the sales stage of a house. Jobs and employment are provided to many people, such as real estate agents and plumbers, after the housing construction is completed. Buying a house is an important decision to live or invest in. For this reason, when buying a house, it should be investigated in detail and the right option should be selected. Therefore, the factors affecting housing prices should be examined in detail to offer them to the users.

Real Estate Price Indices are known as one of the important indicators showing the developments in house prices [44]. The fact that houses are both used for a long time and an investment tool causes the housing market to differ from other markets. These differences can be stated as high cost, continuity, heterogeneity, being fixed, causing growth in secondary markets, and being used as collateral [1,2].

Nowadays, people prefer to buy their own houses by taking long-term mortgage loans rather than being tenants, especially with the emergence of mortgage loans. In periods when interest rates on housing loans decrease, with the increasing demand for housing purchases, homeowners may see this as an opportunity and tend to increase housing prices [3-6]. On the other hand, factors affecting the country's economy are thought to be directly related to housing prices. Therefore, inputs such as gold price, iron price, and minimum wage on the date the house is announced are included in this study [7-9]. New vehicle prices are also a determining factor in the country's economy. So it will be entered in the price of one of Turkey's best-selling cars [10-14].

The impact of universities on urban economies is unquestionably high. A university undergraduate student who comes to a city for education contributes economically to city tradesmen for an average of 4 years. In addition to the economic contribution of a student to sectors such as transportation, food, textile, and entertainment, the most important need is for accommodation. In addition to student dormitories and apartments, most students settle in rented homes alone or with friends. The landlords may be renting out their homes for more than a normal rental price, knowing that the students do not pay the rent alone but share it. This will mean that the houses near the university have high rental income, so it is a very important investment opportunity. As a result, it is expected that houses near the university will be more attractive to buy and increase house prices. Therefore, in this study, by adding the variable of proximity to the university, it is tested whether this variable affects the housing price or not.In addition to the above-mentioned economic factors and proximity to the university, Sakarya province was chosen to test factors such as transportation, tourist places, and proximity to the sea.

There are two state universities in Sakarya, namely Sakarya University and Sakarya Applied Sciences University. The main campus of Sakarya University is in the Serdivan district and it has different faculties in other districts of Sakarya. Karasu district is an important district with a coastline. The high-speed train station is located in Arifiye district. There are train services from Arifiye to cities such as Ankara, Istanbul, and Konya. Adapazarı district is known as the central district and is accepted as the city center. At the same time, there is AdaRay station in the Adapazarı district, which also has train services to Istanbul. For this reason, sample data for this year and previous years will be selected from these districts, as it is thought that it will help to get accurate results in this study. Artificial Neural Networks (ANN), one of the machine learning methods, are used to solve the problem in many different areas [15]. Most of the studies conducted with the ANN method for estimating the house price show that ANN models perform well in estimating house prices [1]. The aim of this study is to determine the best inputs which can be used in house price estimation by comparing deep learning, polynomial regression, and random forest (RF)methods.

BSEU Journal of Science
https://doi.org/10.35193/bseufbd.998331

e-ISSN:2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)

II. LITERATURE REVIEW

ANN have been used in estimating house prices since the early 1990s. Borst's first study in 1991 draws the attention [16]. On the other hand, Do and Grudnitski suggested that in their study with 105 inputs in 1992, the ANN model performed better than the multiple regression model in house price estimation [17].There are many complicated criteria in housing price determination, so care should be taken when choosing and interpreting variables. Three different studies were carried out in the provinces of Aydın, Balıkesir and Giresun in 2021 on the criteria for determining housing prices [18-20]. According to these studies, it has been concluded that the criteria determining housing prices are similar. It is important to reach the resources containing the required inputs and to design the resources and inputs used correctly to achieve a highly consistent result with ANN [1].

While finding solutions to classification problems, as the number of groups and parameters in the data set increases, the models used with ANN begin to perform better than parametric models. In ANN training, three data sets are required: training set, evaluation set, and test set. However, it is not easy to obtain a sufficient sample for this, especially in accounting and financial applications. As a solution to this situation, there are Jackknife and cross-validation methods applied on small data sets. In this study, without dividing the data set into three groups, Jackknife and cross-validation methods were shown to be effective techniques [21].

In Ecer's study in 2014, housing price estimation was made using the data of 610 houses in Izmir. The estimation results of the hedonic model and the Multilayered Perceptron (MLP) model, which is one of the ANN model were compared in the estimation of housing prices. The dataset variables consist of the properties of the houses. These features are price, size, facade, age, location, floor, number of rooms, exterior features of the house such as elevator, swimming pool, interior features such as ADSL, shower cabin, type of house such as garden, duplex. According to the hedonic model, the size of the house, the presence of a shower cabin, dressing room, pantry, in-suite bathroom, walk-in closet, built-in kitchen, blinds, jacuzzi, water heater, elevator, indoor garage, and outdoor swimming pool, proximity to the primary school and the pier, the house has a sea view are the most important variables affecting the price. In the MLP model, the variables affecting the housing price are as follows: Swimming pool, being by the sea, close to the city center, having a burglar alarm, having a built-in closet, age, being close to the train station, having security on the site, being close to the sea bus, being close to the mosque, university and proximity to the health center, being on the street, having a built-in kitchen, sound insulation, insulating glass, cable TV, hydrophore, steel door, and terrace. RMSE, MAE, MAD, and Theil U performance measures were used to compare the prediction accuracies of the hedonic model and the MLP model. These four criteria are based on error values. The values close to zero indicate that the prediction performance of the model used is high. The obtained values from the MLP model are approximately 60-90% smaller than hedonic model. Furthermore, it has been concluded that the MLP model better than the hedonic model. The interest and exchange rates affect the housing price which are among the ergonomic factors, are not included in the forecast models which limits the study [22].

Limsombunchai used the ANN model in the study, taking home prices data from Christchurch, New Zealand. Building age information, building type, residence, residential area, social facilities, location parameters and garage number are used. It is stated that, the accuracy of the model is about 84% [23].

Özkan et al., took 170 house prices in the Selçuklu district of Konya as input and used ANN and regression methods to estimate the prices of houses. The size of the house, age of the building, floor, facade, base area coefficient, floor area coefficient, and location are used as parameters. An accuracy rate of 84.5% was obtained with ANN and 83% with the regression method respectively [24].

Zurada et al., tried to estimate housing prices using fuzzy logic and ANN. They chose parameters as the number of bathrooms, residential area, garage size, heating system, fireplace, construction type, and garage type. As a result, they stated that fuzzy logic and ANN methods which can be used in house price estimation if sufficient data is done and correct [25].

Yılmazel et al., conducted a study using a hedonic model with the data of 5556 houses which were bought from Eskişehir in 2018. They took the neighborhood, where it is located, the size of the house, the number of rooms, the number of bathrooms, the car park, the en-suite bathroom, the kitchen, the elevator, the distance to the tram stops, the number of floors, whether the apartment is on the 1st floor, whether there is central heating, whether there is a parent bathroom, the presence of internet values as parameters. They obtained the most accurate result with a correlation coefficient (r) of 0.9219. This gives the result 0.84989961 from r * r when viewed as r squared.

In this study, they stated that artificial neural network techniques can be used as a tool for estimating house prices [1].

Sarip and Hafez analyses the performance for the estimation of house prices with non-linear models for distinct area in Malaysia. The investigation of the Fuzzy logic model performance in predicting housing prices have been searched. They argued that the obtained MAE values were acceptable with the using of three models which are fuzzy logic, ANN and FIS. To deal of estimating real estate prices, the fuzzy logic based approach is more accurate than other ones in modelling the relationship between real estate price and real estate properties [26]. Afonso and friends are collected 12,223,582 housing advertisements data from Brazilian websites from 2015 to 2018 to estimate housing prices in Brazil in 2019. They applied two different methods (Enriched RF and Recurrent Neural Networks) and finally combined them. In the combined model, the score obtained as a result of the estimation was 0.23847 Root Mean Squared Log Error *(*RMSLE). It was observed that two models created with RF and Recurrent Neural Networks (RNN) gave good results. They found that the Enriched RF worked well with numerical features and gave good results but were unable to process the raw image or text data. On the other hand, RNN observed that it is suitable for all kinds of data but does not process numerical features and RF. They advocated combining these two methods, as both methods have their respective strengths. Experimental results show that strengthening the dataset and combining different machine learning applications outperform for prediction of housing prices in Brazil [27].

Wang and Wu used a total of 27649 data (houses), which included the prices for single-family home evaluation of Virginia USA. They used features such as the size and year of construction that clearly define the quality of the home, but they did not have any specific variables that clearly define home comfort. Using the RF algorithm with location, age, and size of the building, they got the best result with a result of 0.701680 [28]. Tabales et al. tested the ANN approach as an alternative to hedonic modeling to estimate the selling price of an apartment. For this purpose, they used a fairly large sample of data in a medium-sized city in the south of Spain, stating that size is the most important variable, and they concluded that R^2 was 86.05% [29]. Tabar et al. conducted a valuation study for the residences in Tokat in 2021. The size of the houses, the number of rooms, the age of the building, location, floor, number of bathrooms, and the number of balconies were get as parameters. For value determination ANN and Multiple Regression were compared. ANN model accuracy is about the 96% and better than multiple regression [30].

With the big data visualization application, Erkurt and Yıldırım put forward a study on the housing sector in Turkey. Infographics were produced by determining the number of rooms + halls, age of the building, net area, address, price, and currency of the house in a data set containing 79,632 house advertisements belonging to the provinces of Istanbul, Ankara, and Izmir. With these infographics, individuals will be able to examine the advertisements in the provinces and districts they want to compare according to the criteria for choosing a house, without the need for a data set. They will have the chance to see the details of the advertisement that interests them on the visual. Infographics that update themselves with changes in data are presented as a user-friendly tool [31].

III. MATERIAL AND METHOD

From the previous studies mentioned in the literature review, it has been suggested that the ANN method is more successful than linear methods in estimating house prices and more accurate results can be obtained when more data are used. Therefore, the deep learning method with more hidden layers compared with other machine learning methods like polynomial regression and RF algorithm.

A. Dataset

As input, it was stated that positive results were obtained in previous studies [1,23-25], the size of the house (net square meters), number of rooms, ϵ, age of the building, floor, number of floors, heating type, number of bathrooms, having a car park, having a pool, security, whether it is a duplex, whether it has a terrace, whether it has a sea view, whether it is in the city center distance to the university were taken. Besides being close to the train station, being in site, gram price of gold, kilogram price of iron, car price and most importantly, bank housing loan interest rates in that period also were taken.

To test all the criteria, the city of Sakarya, which has a seafront, train stations, and 2 universities, was selected.166 advertisements placed in January 2021, 11 advertised in January 2020, 24 advertised in January 2018, and 12 advertised in February 2016 from Serdivan, Arifiye, Adapazarı, and Karasu districts in Sakarya were

randomly taken from various real estate advertisement sites. For evaluation of the algorithms 0.33 of the data is reserved as test and the rest as training.

B. Machine Learning Algorithms

Machine learning algorithms perform classification and clustering operations using supervised and unsupervised learning styles. Deep learning methods are used in the rapid learning and application of complex data by removing the human factor from machine learning [32]. After the development of technology and the discovery of high hardware products such as GPUs and the amount of data and resources that can be used increased, artificial neural networks started to become popular again by increasing the number of hidden layers and nodes. Thus, deep learning has emerged.

1) Deep Learning:

It is a branch of machine learning. Just as people learn from their experience, deep learning algorithms make changes in their steps to the result each time to get a better result. The algorithms used can be supervised or unsupervised. Deep learning has a nonlinear one or more-layer structure to perform feature extraction and conversion. In the deep learning network, which has a successively layered structure, a separate process is carried out on each layer, and data is left to the next layer. That is, the output of one layer becomes the input of the next layer [32]. Figure 1 shows the deep learning model in Figure 1. As seen in Table 3, by changing the numbers of hidden layers, neurons and epochs experimentally, it was observed that the most accurate result was obtained with an input layer with 21 neurons, 2 intermediate layers with 3 and 11 neurons, and 1 output layer with 350 epochs.

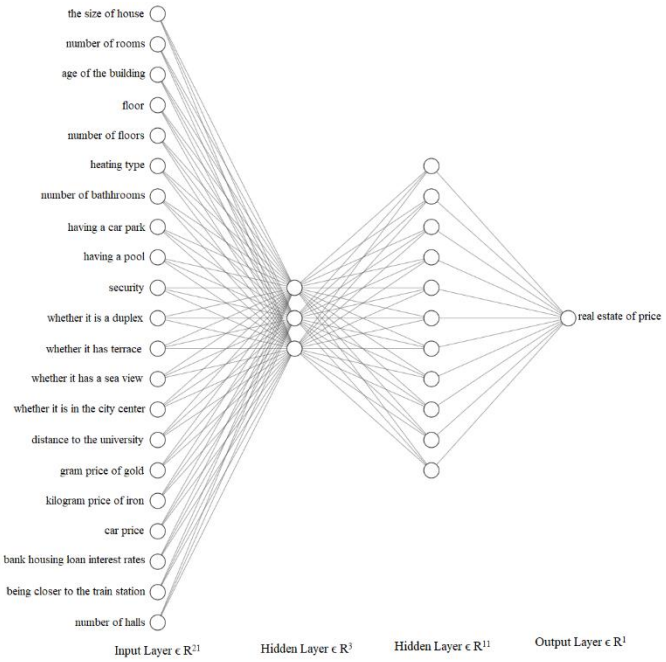

Figure 1. Deep learning model

2) Polynomial Regression:

Regression determines the relationship between variables. One of them is dependent and the other / others are independent variables. The method used to show the relationship is called Regression analysis. With the regression analysis, the results of the dependent variable are interpreted according to the movement of the independent variables. When the relationship between variables is not linear, Polynomial Regression is used as analysis method [35]. Y result value for a single prediction of X is shown in equation 1:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_h X^h + \varepsilon \quad (1)$$

In equation 1, h is called the polynomial degree. The relationship between variables for low degrees of h is called as h = 2 quadratic, h = 3 cubic, etc. Equation 3 expresses both polynomial regression and linear regression.

Because there is a nonlinear relationship between X and Y variables. The coefficients of β_1, β_2,…, β_h are also linear [35].

3) Random Forest Algorithm:

Classification algorithms are used to predict which class of data will be included. More than one classifier is used in ensemble approaches. These classifiers are combined such as voting, average and different accuracy rates in process. With the help of this it offers better results than single classifier. One of the most widely used batch classifiers is the RF which consists of multiple decision trees. This algorithm is used for classification and regression. RF has emerged by improving techniques in Ensemble methods and adding the feature of randomness [36].

Training data set to be used for RF is allocated as N sub samples. Each sub sample is trained in a decision tree. Finally, a trained model with N decision trees is formed. All decision trees in the RF work for the classification of an incoming sample. The average of each working model result is taken and the final estimation is made and the incoming sample is classified.

The trees in the model are large and different from each other. Therefore, the correlation between trees is not high and biased results are not produced. Since the results are unbiased, the accuracy value of the model created with RF among machine learning methods is higher and the model is reliable [37].

The prediction of the kth tree for a given input x is given as shown in equation 2[27].K in the equation shows the number of trees. Θ_k is the random vector equivalent of each sub-samples that will be trained with the decision tree.

$$h_k(x) = h(x, \Theta_k), \forall k \in \{1,2,\ldots,K\} \quad (2)$$

Node S can be split into two sub-nodes in multiple ways. Assuming that a threshold c is chosen for the selected feature by dividing S by S1, S2 concerning each feature value vi. A c can be used for a regression task that minimizes the difference in the sum of error squares (SSE) as shown in equation 3 [27].

$$SSE = \left(\sum_{i\in S1}(v_i - \frac{1}{|S_1|}\sum_{i\in S1} v_i)^2 + \sum_{i\in S2}(v_i - \frac{1}{|S_2|}\sum_{i\in S2} v_i)^2\right) \quad (3)$$

The final estimate to be made with the random forest is found by taking the average of the estimation result of all decision trees as shown in equation 4 [27].

$$h(x) = \frac{1}{K}\sum_{i=1}^{K} h_k(x) \quad (4)$$

C. Activation Functions

In the activation function, the neural network cell obtains transaction output results on inputs. It is important to choose this activation function correctly. Because it will affect the result and result performance. The function can be a single and double pole [33]. There are various activation functions such as Sigmoid, Rectified Linear Unit (RELU), Step function.

1) Sigmoid Function:

The dynamic range of change of the sigmoid activation function is the range [0 1], and the function exhibits a nonlinear change in this range [34]. Equation 1 shows the input-output expression of the sigmoid function, where n is a real number between -∞ and + ∞ and e is the Euler number, and the change of the function concerning the input is shown in equation 5.

$$A(n) = \frac{1}{1+e^{-n}} \quad (5)$$

2) Rectified Linear Unit (RELU) Function:

It is a non-linear function. RELU function takes the value 0 fo negative k inputs, k value for positive k inputs. RELU formula for k positive values is as shown in equation 6.

$$R(k)=max(0,k) \quad (6)$$

D. Variable Selection Methods for Machine Learning Models

1) Least Squares Method (LSA):

The LSA is a standard regression method used to write down the relationship between two dependently varying physical quantities as an equation that is as realistic as possible. Furthermore, this method serves to find a function curve that will pass as close as possible to the measured data points. Linear least squares is estimating unknown parameters in a linear regression model which aims to minimize the sum of the squares of the variables among the observed responses. It is compared with data predicted by a linear function of a set of explanatory variables. It is viewed as the sum of the squared vertical distances between the corresponding point on the regression line and each data point in the set. If the differences are small, the model should fit the data better [38].

2) P-Value (Probability Value):

P-value is the probability that the relationship or difference between variables used in a sample has arisen accidentally, although there is no relationship or difference in reality [39].

Taking a value between 0 and 1 and a P-value of 0 means that the relationship between the variables is the strongest, and a 1 means the weakest.

E. Performance Measurement Metrics for Machine Learning Models

1) Correlation Coefficient (R^2 Value):

The R^2 value is a measure that represents predictive performance. Explains the percentage volume of cases where the independent variable affects the dependent variable, the capacity of the relationship between the independent and dependent variable is shown in the range of 0-100%. It can be said that as the R^2 value approaches 100%, the relationship is very strong and the accuracy of the predictions to be made through the model will also show a good performance. The low values of the metrics indicate that the standard deviation is high [40]. Correlation Coefficient as shown in equation 7. R is a correlation coefficient, n shows the number in the given dataset, y is a first variable in context and y' second variable.

$$R^2 = \left(\frac{n\sum y.y' - (\sum y)(\sum y\prime)}{\sqrt{n(\sum y^2) - (\sum y)^2} - \sqrt{n(\sum y\prime^2) - (\sum y\prime)^2}}\right)^2 \quad (7)$$

2) Mean Squared Error (MSE):

The MSE represents how close the regression curve is to a set of points. By using it as estimator which measures the performance of the machine learning model. The value of it, always positive then it can be said that estimators with MSE value close to zero perform better [43] which can be obtained in equation 8. x is the property the model uses to make the prediction and prediction(x) meaning is the predicted value based on the x. y represents the actual value and N is the number of samples. In case of [42]:

$$MSE = \frac{1}{N} * \sum_{x,y \in D} y - (prediction(x))^2 \quad (8)$$

3) Root Mean Squared Error (RMSE):

The magnitude of error of a machine learning model is measured by it which is a quadratic metric. It is often used to find the distance between the predicted values of the estimator and the true values. It shows how much the data is around the line. It takes values between 0 to ∞. At the lower values estimators perform better. When the RMSE value equals to zero which means that the model made no errors [41]. RMSE values is calculated as equation 9.

$$RMSE = \sqrt{MSE} \quad (9)$$

IV. RESULTS AND DISCUSSIONS

Records are obtained from various house advertisements from Serdivan, Adapazarı, Arifiye and Karasu districts of Sakarya province in February 2016, January 2018, January 2020, and January 2021, consist of 213 advertisements as in Table 1, and the most expensive house is 520,000 TL, the cheapest house is 90.000 TL and an average price of houses is 262.000 TL.

Table 1. House prices in the dataset

Number of Residences	Maximum Price (TL)	Minimum Price (TL)	AveragePrice (TL)
213	520.000	90.000	262.000

To estimate the prices of the houses in the data set, the inputs mentioned in the dataset used section were given to the system by using Deep learning, Polynomial Regression, and Random Forest algorithms through the python programming language, and outputs were obtained. For deep learning, firstly the inputs were scaled with the Min Max Normalization method and converted to 0-1 interval like in equation 10.

$$h(x) = \frac{Xi - \min(x)}{\max(x) - \min(x)} \tag{10}$$

It was first determined that the number of living rooms criteria were the same in all inputs thus, this feature was chosen as the first variable to be excluded in the estimations. As the inputs were removed in order, p values were obtained. The backward elimination method and the stats model library in the python programming language were used to calculate the p values [41]. This library firstly takes all the independent variables as input, then estimates the dependent variable with the Ordinary Least Squares method and tests all the inputs and gives the p values of the effects of the inputs on the result. In this study, first of all, all the inputs were given as independent variables and the price variable as the dependent variable, and the first p values were calculated. Then, the variable with the highest p value (Being in the site = 0.906) was removed from the inputs with the backward elimination method and new p values were calculated again. The same iteration was repeated 7 times and the p values were determined as shown in Table 2.

Table 2. Obtained P-Values

Variable	P Value
Being in the Site	0,906
Number of Bathrooms	0,880
Security	0,927
Railway Station	0,819
Iron Price (kg)	0,730
Interest Rate	0,570
Having a Pool	0,576

In order to make predictions with deep learning, the variables were extracted one by one with the python Keras library, and the number of layers and epochs were changed experimentally. The obtained results were compared as in Table 3. RELU function is used in the input and hidden layers, and the sigmoid function is used as the activation function in the output layer.

Table 3. Deep learning prediction results

Subtracted Variables	Number of Input Layer Neurons	Number of Hidden Layer Neurons	Number of Hidden Layer Neurons	Number of Output Layer Neurons	Number of Epoch	R Square
None	23	3	12	1	240	0.7738
None	23	3	12	1	300	0.8460
None	23	3	12	1	350	0.8093
None	23	3	12	1	375	0.7975
Number of living rooms	22	3	12	1	300	0.8383
Number of living rooms	22	3	12	1	240	0.7622
Number of living rooms	22	3	11	1	300	0.7953
Number of living rooms, In Site	21	3	11	1	300	0.8348
Number of living rooms, In Site	21	3	11	1	354	0.8428
Number of living rooms, In Site	21	3	11	1	350	**0.8614**
Number of living rooms, In Site	21	3	11	1	400	0.8060
Number of living rooms, In Site	21	2	11	1	350	0.8540
Number of living rooms, In Site	21	1	11	1	350	0.8498
Number of living rooms, In Site	21	4	11	1	350	0.8244
Number of living rooms, inside the site, number of bathrooms	20	3	11	1	300	0.8390
Number of living rooms, inside the site, number of bathrooms	20	3	11	1	350	0.8219
Number of living rooms, inside the site, number of bathrooms, security	19	3	10	1	350	0.8119
Number of living rooms, inside the site, number of bathrooms, security, proximity to the train station, iron price	17	3	9	1	350	0.8314

The best result with deep learning was obtained by choosing 0.8614 r square, 0.01199 MSE and 0.10948 RMSE and the number of living rooms and the in-site variable, the number of input layer neurons was 21, the number of hidden layers was 3, the number of interlayer neurons was 11, and the number of output layer neurons was selected with 350 epochs. The best results obtained with the predictions made with Polynomial Regression and Random Forest, which are other machine learning techniques, are shown in Table 4.

Table 4. Polynomial regression and random forest prediction results

Method Used	Subtracted Variables	R Square	MSE	RMSE
Polynomial Regression	Number of halls, being in the site, number of bathrooms, security, proximity to the train station, iron price(kg)	0.7651	0.01263	0.11240
Random Forest	Number of living rooms, being in the site, number of bathrooms, security	0.7719	0.01258	0.11214
The Deep Learning Model, which we obtained the best results in this study,	Number of living rooms, In Site	**0.8614**	0.01199	0.10948

The best result from polynomial regression and RF methods was obtained when the variables of number of halls, in-site, number of bathrooms, and safety were removed with 0.7719 r square. However, this result was not more successful than the deep learning method, which gave the best result with 0.8614. The required results corresponding to the entries are shown as lines, and the estimates are shown as dots in the in Figure 2, 3 and 4 respectively.

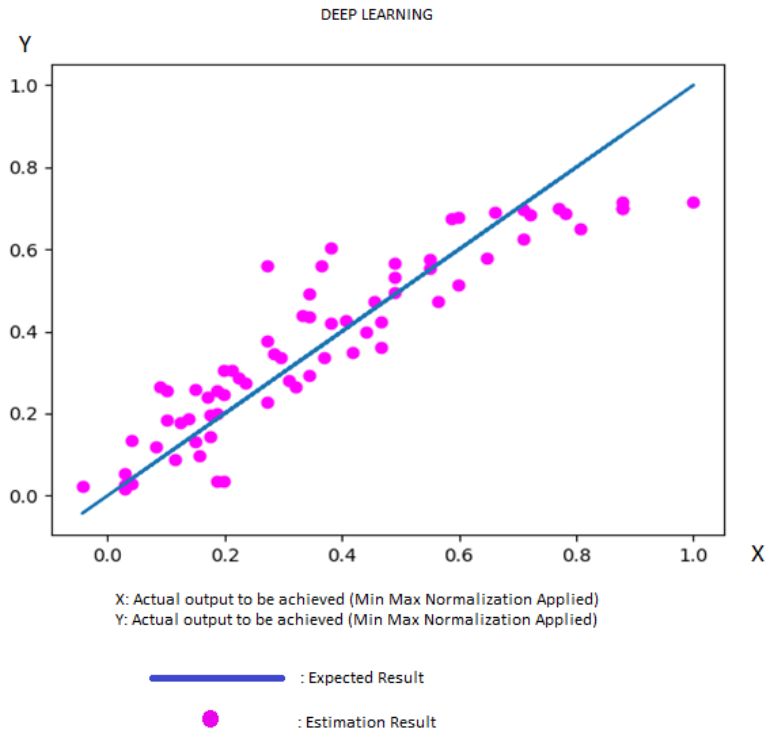

Figure 2. Prediction results with deep learning

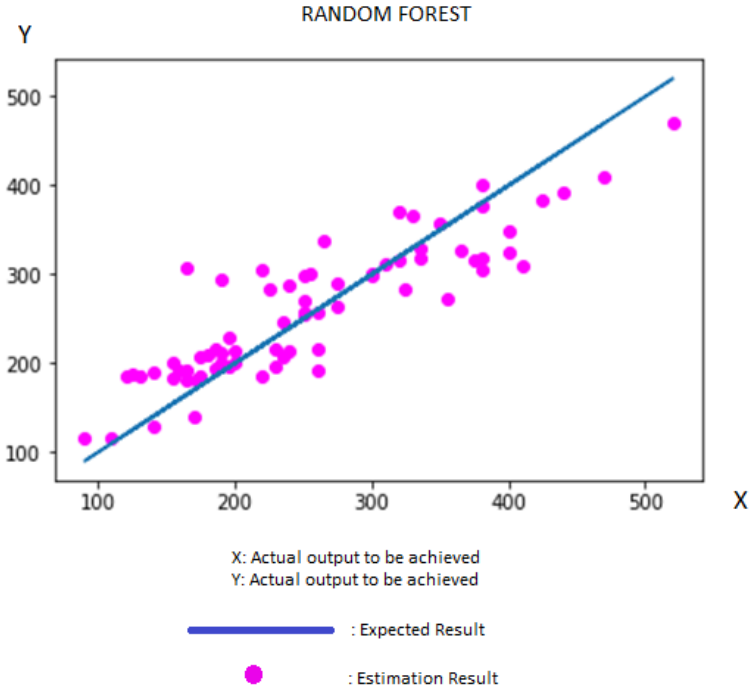

Figure 3. Prediction results with random forest

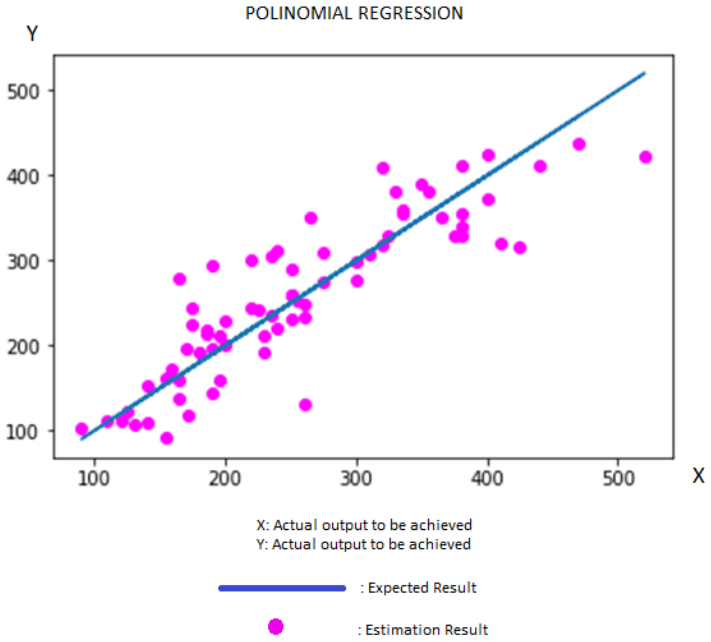

Figure 4. Prediction results with polynomial regression

Furthermore, it was taken that the most successful algorithm among deep learning, polynomial regression, and random forest algorithms in house price estimation is the deep learning algorithm with 86.14% as seen in Table 5.

Table 5. Comparison with other studies

Work	Year	Method	Success
Limsombunchai [23]	2004	ANN	%84
Özkan et al. [24]	2007	ANN	%84,5
Yılmazel et al. [1]	2017	ANN	%84,99
Afonso et al. [27]	2019	Enriched RF and KISS	0.23847 RMSLE
Wang and Wu [28]	2018	Random Forest	%70,17
Proposed Method	2021	Deep Learning	%86,14

V. CONCLUSION

The size of the house, the number of rooms, the age of the building, the number of floors, the type of heating, the number of bathrooms, parking, pool, security, duplex, terrace, sea view and city center, proximity to the university determine the housing prices. Also, it was understood that gram gold price, price per kg of iron, new vehicle price, and bank housing loan interest rates were the economic factors of the period. In this study, the number of halls criterion is ineffective since all residences have a single hall. Also, it was understood that being on the site did not affect the price of the house, and it was understood that the features such as pool, parking lot, and security on the site were more effective in price estimation.

The best second value obtained with RF algorithm with 77.19% as seen in Table 4. The number of halls and the presence in the site, as well as the number of bathrooms and the presence of security in the site parameters gave the best result.

It was seen that economic factors, such as gram gold price, iron kg price, new vehicle price, bank housing loan interest rates have affected the house price, and since there are long day differences between the dates selected together, their effect on housing prices may be more apparent. Therefore, to determine these factors more accurately, it can be suggested to test the effect of the gram gold price on the house prices when the price of gold rose and fell in the same short period. At the same time, increasing the number of data as the factor to be recommended for testing can be offered for accurate result detection.

REFERENCES

[1] Yılmazel, Ö., Afşar, A., & Yılmazel, S. (2018). Konut fiyat tahmininde yapay sinir ağları yönteminin kullanılması. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, *20*, 285-300.

[2] Lacovıello, M. (2000). House prices and the macroeconomy in Europe: results from a structural VAR analysis", ECB Working paper.

[3] Para ve Borsa. (2018). *Konut Kredisi Faiz Oranları,*https://www.paraborsa.net/i/ocak-2018-konut-kredisi-faiz-oranlari/, (17.02.2021).

[4] Mynet (2021). *Bankaların kredi faiz oranları, Ziraat, Vakıfbank, Akbank, TEB, İş ve Halkbank ihtiyaç, taşıt ve konut kredisi faiz oranları,* https://finans.mynet.com/haber/detay/ekonomi/bankalarin-kredi-faiz-oranlari-27-ocak-2021-ziraat-vakifbank-akbank-teb-is-ve-halkbank-ihtiyac-tasit-ve-konut-kredisi-faiz-oranlari/413494/, (17.02.2021).

[5] Emlak Kulisi, (2021). *Konut kredisi faiz oranlarında üç bankada yükseliş.*https://emlakkulisi.com/konut-kredisi-faiz-oranlarinda-uc-bankada-yukselis/450201, (17.02.2021).

[6] Takvim, (2020). *Ziraat Bankası konut, ihtiyaç ve taşıt kredisi faiz oranları*. https://www.takvim.com.tr/ekonomi/2020/01/30/049-ve-079-ziraat-bankasi-kredi-faiz-oranlari-firsati-31-ocak-2020-konut-ihtiyac-ve-tasit-kredisi-faiz-oranlari, (17.02.2021).

[7] Bigpara, (2021). *Altın*. https://bigpara.hurriyet.com.tr/altin/, (17.02.2021).

[8] Demir Fiyatları, (2021). *Demir Fiyatlari*,https://www.demirfiyatlari.com/, (17.02.2021).

[9] iKMagazin, (2021). *Yıllara Göre Asgari Ücretler Tablosu*, https://ikmagazin.com/insan-kaynaklari/yillara-gore-asgari-ucretler-tablosu/, (17.02.2021).
[10] NTV, *2020'nin en çok satan araba modelleri (Hangi otomobil markası kaç adet sattı?),* https://www.ntv.com.tr/galeri/otomobil/2020nin-en-cok-satan-araba-modelleri-hangi-otomobil-markasi-kac-adet-satti,72hIZ6U6fEaQLr4okdn6KA/FGCOLEZgVU61F4aaL4LV9g, (06.02.2021).
[11] Yeni Model Arabalar, *Fiat Ocak Kampanyası, Fiyat Listesi - Ocak 2020-01-06*, https://www.yenimodelarabalar.com/fiat-ocak-kampanyasi-fiyat-listesi-ocak-2020-01-06-39045.html, (17.02.2021).
[12] Otopark, *Fiat Şubat 2016 Fiyat Listesi*, https://otopark.com/2016/02/08/fiat-subat-2016-fiyat-listesi/, (17.02.2021).
[13] Yeni Model Arabalar, *2021 Fiat Egea Fiyat Listesi – Özellikleri - Ocak 2021-01-04*, https://www.yenimodelarabalar.com/2021-fiat-egea-fiyat-listesi-ozellikleri-ocak-2021-01-04-41595.html, (17.02.2021).
[14] Otomobilblog.com, (2018). *2018 Ocak Ayı Fiat Otomobil Fiyat Listesi*, https://www.otomobilblog.com/makale/2018-ocak-ayi-fiat-otomobil-fiyat-listesi_852, (17.02.2021).
[15] Çavuşlu, M. A. (2021). Plaka Bölgesi Tespiti Problemi için Yapay Arı Koloni Algoritması ile YSA Eğitiminin APKD'de Gerçeklenmesi. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, *8*(1), 446-457.
[16] Borst, R.A. (1991). Artificial neural networks: the next modelling/calibration technology for the assessment community. *Property Tax Journal*, *10.1*, 69-94.
[17] Do, A.Q. & Grudnıtskı, G. (1992). A neural network approach to residential property appraisal. *The Real Estate Appraiser*, *58.3*, 38-45.
[18] Aydoğdu, N. (2021). *Bulanık regresyon analizi ile Aydın ili konut fiyatlarını etkileyen değişkenlerin belirlenmesi*, Master's thesis, Aydın Adnan Menderes Üniversitesi Fen Bilimleri Enstitüsü.
[19] Yildirim, H. H. (2021). Balikesir İlinde Konut Fiyatlarindaki Farklılıkları Oluşturan Unsurların Tespiti. *Uluslararası Finansal Ekonomi ve Bankacılık Uygulamaları Dergisi*, *2*(1), 38-62.
[20] Yeşil, P., & Güzel, M. (2021). Giresun kent merkezi'nde konut fiyatlarına etkieden yapısal ve çevresel etkenlerin belirlenmesi. *Akademik Ziraat Dergisi*, *10*(2), 305-316.
[21] Coakley, J.R. & Brown, C.E. (2000). Artificial neural networks in accounting and finance: Modeling issues. *Intelligent Systems in Accounting, Finance & Management*, *9(2)*, 119-144.
[22] Ecer, F. (2014). Türkiye'deki konut fiyatlarının tahmininde hedonic regresyon yöntemi ile yapay sinir ağlarının karşılaştırılması. *In International Conference On Eurasian Economies*, 1-3 July 2014 , Skopje, 1-10.
[23] Limsombunchai, V. (2004). House price prediction: hedonic price model vs. artificial neural network.*New Zealand agricultural and resource economics society conference*, 25-26 June 2004, Blenheim, 25-26.
[24] Özkan, G. Yalpır, Ş. & Uygunol, O. (2007). An investigation on the price estimation of residable real-estates by using artificial neural network and regression methods. *12th Applied Stochastic Models and Data Analysis International conference (ASMDA)*, 29May-1 Junr2007 ,Chania, Crete, Greece,1-8.
[25] Zurada, J. M. Levitan, A.S. & Guan, J. (2006). Non-conventional approaches to property value assessment. *Journal of Applied Business Research (JABR)*, 22(3).
[26] Sarıp, A. G. & Hafez, M.B. (2015). Fuzzy logic application for house price prediction. *International Journal of Property Sciences (E-ISSN: 2229-8568)*, 5.1.
[27] Afonso, B. K. Luckeciano, C. M. Willian, D. Samuel, S. & Berton, B. (2019). Housing Prices Prediction with a Deep Learning and Random Forest Ensemble. *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional. SBC*, 389-400.
[28] Wang, C. & Wu H. (2018). A new machine learning approach to house price estimation. *New Trends in Mathematical Sciences*, *6(4)*, 165-171.
[29] Tabales, J.M.N., Caridad, J.M., & Carmona, F. J. R. (2013). Artificial neural networks for predicting real estate price, *Revista de Métodos Cuantitativos para la Economía y la Empresa*, *15*, 29-44.

[30] Tabar, M.E., Başara, A. C., & Şişman, Y. (2021). Çoklu Regresyon ve Yapay Sinir Ağları ile Tokat İlinde Konut Değerleme Çalışması. *Türkiye Arazi Yönetimi Dergisi, 3(1),* 1-7.

[31] Erkurt, E. & Yıldırım, E. (2021). Bir büyük very görselleştirme uygulaması olarak konut tercih info grafikleri. *Afyon Kocatepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 23(1),* 36-52.

[32] Şeker A., Diri, B. & Balık, H.H. (2017). Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme. *Gazi Mühendislik Bilimleri Dergisi (GMBD), 3.3,* 47-64.

[33] Öztürk, K. & Şahin, M.E. (2018). Yapay sinir ağları ve yapay zekâ'ya genel bir bakış. *Takvim-I Vekayi, 6(2),* 25-36.

[34] Altun, H., Eminoğlu, U., & Tezekici, B.S. (2014*). MLP Yapay Sinir Ağlarında Öğrenme Sürecinin Aktivasyon Fonksiyonu ve İstatiksel Değişim Gösteren Giriş Verilerine Bağımlılığı.* Mühendislik-Mimarlık Fakültesi Elektrik-Elektronik Mühendisliği Bölümü, Niğde Üniversitesi, *https://www.emo.org.tr/ekler/490c742cd8318b8_ek.pdf, (20.09.2021).*

[35] Veri Bilimcisi, *Polinomsal Regresyon*. https://veribilimcisi.com/2017/07/18/polinomsal-regresyon-polynomial-regression/. (06.02.2021).

[36] Akar, Ö., & Güngör, O. (2012). Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması. *TMMOB Harita ve Kadastro Mühendisleri Odası Jeodezi ve Jeoinformasyon Dergisi, 1(2),* 139-146.

[37] Yılmaz, H. (2014). *Random Forests yönteminde kayıp very probleminin incelenmesi ve sağlık alanında bir Uygulama*. Master's Thesis, Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü, Eskişehir.

[38] Veri Bilimcisi, *Polinomsal Regresyon*. https://veribilimcisi.com/2017/07/13/siradan-en-kucuk-kareler-yontemi/.(13.06.2021).

[39] Biistatistik, *P Değeri*. https://biistatistik.com/p-degeri-caninizi-sikmasin/.(02.02.2021).

[40] TR Education, *r squared formula*.https://tr.education-kit.net/7582646-r-squared-formula. (14.06.2021).

[41] Veri Bilimcisi, *RMSE-MAE-MAPE Metrikleri*, https://veribilimcisi.com/2017/07/14/mse-rmse-mae-mape-metrikleri-nedir/. (14.06.2021).

[42] Medium, Model Performans Metrikleri. https://medium.com/deep-learning-turkiye/model-performans%C4%B1n%C4%B1-de%C4%9Ferlendirmek-metrikler-cb6568705b1. (14.06.2021).

[43] Statsmodel library, *Statsmodel Library*. https://www.statsmodels.org/stable/index.html. (13.06.2021).

[44] Savuran, O. İ. (2008). *Gayrimenkul fiyat endeksleri ve İstanbul Etiler bölgesinde konutlar için hedonic fiyat endeksi uygulaması*. Master's Thesis, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü , İstanbul.