REPUBLIC OF TURKEY

ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DEPARTMENT OF INDUSTRIAL ENGINEERING

HYBRID SOFT COMPUTING METHODS FOR IMPROVING REAL ESTATE PRICE FORECASTING

NURAN MEMİLİ

MASTER OF SCIENCE

ADANA 2019

REPUBLIC OF TURKEY
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF INDUSTRIAL ENGINEERING

HYBRID SOFT COMPUTING METHODS FOR IMPROVING REAL ESTATE PRICE FORECASTING

NURAN MEMİLİ
MASTER OF SCIENCE

SUPERVISOR
Assoc. Prof. Dr. MUSTAFA GÖÇKEN

ADANA 2019

I hereby declare that all information in this thesis has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all information that is not original to this work.

Nuran MEMİLİ

ABSTRACT

HYBRID SOFT COMPUTING METHODS FOR IMPROVING REAL ESTATE PRICE FORECASTING

Nuran MEMİLİ
Department of Industrial Engineering
Supervisor: Assoc. Prof. Dr. Mustafa GÖÇKEN
09 2019, 91 pages

Real estate has a very important place for countries and the world economy. The fact that real estate has a heterogeneous structure, that is, the diversity of the properties that make up each property itself, shows that the methods to be selected for determining the sale and rent value of the real estate are important. For real estate price estimation, it can be said that non-traditional methods are more successful than traditional ones. In recent years, metaheuristic approaches for real estate valuation have increased and researches on this subject are continuing. The superiority of metaheuristics in solving non-linear and complex problems provides an important advantage for real estate valuation. In this study, a hybrid approach has been developed for real estate valuation by using artificial neural networks and feature selection methods. The grid search method was used for the most suitable model parameters before the analysis with the artificial neural network. It is aimed to increase the model success by optimizing the parameters that can be selected for the model with grid search method. In addition, in order to minimize possible errors and prevent overfitting during training, the data was divided into appropriate training sets by cross validation technique. The selected data set for this study contains housing data of various districts of Istanbul province, which has a significant value for Turkey. Considering that data preprocessing is important for the success of the model, data preprocessing steps were performed for each district data. After the data preprocessing process, firstly feature selection method, then grid search method with cross validation and finally artificial neural network method with cross validation were applied step by step. The advantage and success of the artificial neural network method, which is one of

the metaheuristic approaches, is supported by this study for real estate price estimation involving many variables. With this model developed for real estate appraisal, it is aimed to examine the properties that affect real estate prices and to contribute to valuation methods in addition to finding realistic price estimation. In addition, it is an important advantage that it is possible to follow the price changes in the real estate market which has a heterogeneous structure with this research, and it is thought that it will provide benefit for the subsequent real estate appraisal studies.

Keywords: real estate appraisal, artificial neural networks, k-fold cross validation, data preprocessing

ÖZET

GAYRİMENKUL FİYAT TAHMİNİNİ İYİLEŞTİRMEK İÇİN HİBRİT YUMUŞAK HESAPLAMA YÖNTEMLERİ

Nuran MEMİLİ

Endüstri Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Mustafa GÖÇKEN

09, 2019, 91 sayfa

Gayrimenkul ülke ve dünya ekonomisi için çok önemli bir yere sahiptir. Gayrimenkulün heterojen bir yapıya sahip olması, yani her bir gayrimenkulün kendisini oluşturan özelliklerin çeşitlilik içermesi, gayrimenkulün satış/kira değerinin belirlenmesi için seçilecek yöntemlerin önemli olduğunu göstermektedir. Gayrimenkul fiyat tahmini için, geleneksel olmayan yöntemlerin geleneksel yöntemlere göre daha başarılı olduğu söylenebilir. Son yıllarda gayrimenkul değerlemesine meta-sezgisel yöntemlerle yaklaşımlar artmış olup, bu konu üzerine araştırmalar devam etmektedir. Meta-sezgisel yöntemlerin doğrusal olmayan karmaşık yapıda olan problemlerin çözümündeki üstünlüğü gayrimenkul değerlemesi için avantaj sağlamaktadır. Bu çalışmada gayrimenkul değerlemesi için, Yapay sinir ağları ve özellik seçim yöntemleri kullanılarak hybrid bir yaklaşım geliştirilmiştir. Yapay sinir ağı ile yapılacak analizlerden önce en uygun model parametreleri için ızgara arama yöntemi kullanılmıştır. Izgara arama yöntemi ile model için seçilebilecek parametreler optimize edilerek, başarının artması amaçlanmıştır. Ayrıca olası hataları minimum düzeye indirmek ve eğitim sırasında ezberlemenin önüne geçebilmek için, çapraz doğrulama tekniği ile veriler uygun eğitim setlerine bölünmüştür. Çalışmada Türkiye için önemli bir değere sahip olan İstanbul ilinin çeşitli ilçelerindeki konutların bilgileri kullanılmıştır. Veri ön işlemenin modelin başarısı için önemli olduğu göz önünde bulundurularak, her bir ilçe veriseti için veri ön işleme adımları gerçekleştirilmiştir. Veri ön işleme sürecinden sonra, veriler üzerinde ilk olarak özellik seçim yöntemi, daha sonra çapraz doğrulama ile birlikte ızgara arama yöntemi ve son olarak çapraz doğrulama ile birlikte yapay sinir ağı yöntemi adım adım uygulanmıştır. Çok fazla değişken içeren gayrimenkul fiyat tahmini için, meta-sezgisel yaklaşımlardan yapay

sinir ağı yönteminin avantajı ve başarısı bu çalışma ile desteklenmektedir. Gayrimenkul değerlemesi için geliştirilen bu model ile, gerçeğe yakın fiyat tahminlerinin bulunmasının yanısıra, gayrimenkül fiyatlarını etkileyen özelliklerin de incelenmesi ve bu anlamda değerleme yöntemlerine katkı sağlaması amaçlanmıştır. Önerilen modellerin en önemli avantajı, heterojen bir yapıya sahip olan gayrimenkul piyasasındaki fiyat değişimlerini takip edebilmeyi mümkün kılması olup, bu avantajın sonraki gayrimenkul değerleme çalışmaları için fayda sağlayacağı düşünülmektedir.

Anahtar Kelimeler : gayrimenkül değerlemesi, yapay sinir ağları, k-katlamalı çapraz doğrulama, veri ön işleme

ACKNOWLEDGEMENTS

I would like to thank my supervisor Assoc. Prof. Dr. Mustafa Göçken for his support. I am grateful to Research Assistant Aslı BORU who took the time during my study and always tried to help me in this field.

I would like to thank my family who has always been with me and supported me and increased my faith. I would like to express my special thanks to my dear mother who has always motivated me throughout my working life.

I would like to thank Elif Akar for her faith in me and her support.

TABLES OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
GA	Genetic Algorithm
LRA	Linear Regression Analysis
GA-Ridge	Ridge Regression Coupled with Genetic Algorithm
GA-LRA	Genetic Algorithm with Linear Regression Analysis
G-SVM	Genetic Algorithm and Support Vector Machines
AMR	Adaptive Mutation Rate
GM	Grey Model
MLR	Multiple Linear Regression
MRA	Multiple Regression Analysis
ANFIS	Adaptive Network-Based Fuzzy Inference
ANFIS-GP	Adaptive Network-Based Fuzzy Inference with Grid Partition
ANFIS-SC	Adaptive Network-Based Fuzzy Inference with Sub Clustering
FLSR	Fuzzy Least-Squares Regression
FRBS	Fuzzy Rule Based System
FAHP	Fuzzy Analytic Hierarchy Process
R^2	R-squared
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
MSE	Mean Squared Error
HIS	Home Sales Index
HRI	Home Rental Index
CBRT	Central Bank of the Republic of Turkey
TSI	Turkish Statistical Institute
MLP	Multi-Layer Perceptron
ADALINE	Adaptive Linear Neuron
SQL	Structured Query Language

1. INTRODUCTION

The real estate sector is defined as a safe, healthy, high and multi-faceted speculative structure that meets socio-cultural needs in the eyes of customers. It closely affects factors such as economic growth, development, and welfare of the country. The fact that this sector has an effective share on global financial markets, developments in financial stability, inflation rate, balance of payments, state budget and external economy, is an indication that it is the key to the national economy. When the concept of real estate is evaluated in terms of investment, it is more liable to financial backed securities (capital market) such as treasury bills and government bonds. It is considered to be high return in the long period, a sustainable and guaranteed investment tool for people (Küçükaslan, N., 2015).

Real estate is a concept associated with the process of need as well as investment, and all forms of need can be physiological and psychological as it was elaborated on Maslow's Requirements Hierarchy theory years ago. Maslow, who sees human instincts in order, argues that the most important factor that changes human behavior is the need for something according to the information obtained from his experience. Finding shelter is one of the most important needs of people (McLeod, S. 2007). One dimension of real estate marketing is the timely determination of consumers' needs. Real estate, which is a physical product, is in the category of primary needs. Therefore, which housing types correspond to which requirements needs to be well determined. Such as; only those who need shelter, they assert that the house needs to be more durable and useful. Those who want to meet their investment needs are more interested in the financial aspect of the house.

Real estate types differ in their structure. The characteristics of the real estate sector are classified by taking into consideration factors such as consumers' expectations from the market, changes in purchasing trends, legal regulations, investment opportunities, sales methods and competition environments. Among the types of real estate; there are also residences and commercial properties. The residential market has an active role in the real estate sector and its heterogeneity physically and geographically makes a house price forecast difficult. Therefore, the choice of a prediction model is very important, especially if it is desired to detect the complicated structure of the relation between the related factors with the

requisition. Another important consideration in estimating housing sales prices is that sellers and buyers need this information correctly during sales. At the same time, improving a housing price forecasting model will help estimate future real estate prices and determine real estate policies. The importance of estimating real estate price inflation is reflected in recent research showing that property prices help to predict inflation and its output (Das et al., 2009a; Forni et al., 2003; Stock et al., 2003). Therefore, a true and accurate estimation of real estate prices not only provides policy-makers with valuable information, but also allows them to better control inflation and design policies. Previous studies represent experimental findings confirming the important relationships between the real estate price and economic factors such as labor-related factors, revenue, interest rate, costs arising from construction (Tsatsaronis et al., 2004; Wheaton, 1999; Linneman, 1986). However, the measurement of these relationships is not sufficient to estimate a house price. House prices are forecasted to interact with a set of factors as an important sign of inflation and its output. Shortly, several economic factors are helping to predict the increase in real housing prices.

The enormous size of real estate stocks and various profiles have led to the use of computer technologies in real estate appraisal processes. Therefore, applications to facilitate the work of evaluators can be obtained commercially or in-house projects, or through systems developed jointly by users and software developers (Kettani and Oral, 2015). Each subsequent work stressed real estate values such as building location, residential area, distance to the center and developed forecastability of the models. Hence, a house price estimation model not only causes to fill a significant knowledge gap, but also increases the yield of the housing market (Calhoun, 2003).

Recent researches that determine the price of housing by using the latest computer technologies are ongoing. When estimating real estate, a large number of attributes are generally used to identify the actual value of an estate. Traditional approaches can be said to be weaker in estimating housing prices because it is not easy to determine many factors and calculate their effect when making an estimation. Many approaches have been developed to analyze real estate valuation criteria and identifying relevant factors and revealing the best combinations accordingly. The most successful ones are stochastic methods and artificial intelligence methods. With these approaches, accurate results can be reached more quickly (Özkanet et al., 2005).

In this section, the factors affecting the value of real estate are included, in section 1.1. Section 1.2 contains information about the real estate market in Turkey. The information about the housing market is given Section 1.3. The purpose and contributions of this research are reported in Section 1.4.

1.1 Factors Affecting The Value Of The Real Estate

The value of a house is determined by its location, quality and the needs of people. This value increases or decreases over time according to factors affecting the value of the house. The real estate value may increase depending on the result of not being able to produce housing or workplace to meet the need. Otherwise, values may decrease due to excess supply. Setting the value of a real estate objectively and ensuring the credibility of its result can be determined by setting all parameters that affect the real estate value and by linking these determined parameters to a mathematical expression.

The significant factors affecting the value of the house can be listed as follows (Saraç, 2012);

1) Number of rooms (The number of rooms has a direct impact on the value of the house.)
2) Apartment area (The usage area of the house directly affects the value.)
3) Net area (The total area of the room, hall, hallway and balcony means the net area. Exceeding the net field increases the value.)
4) Location (The district and neighborhood where the building is located, affects the value of the real estate.)
5) Social facilities (Sport areas and meeting rooms add value to housing.)
6) Elevator; Pool; Car park (These features increase the value of the house.)
7) Transportation (Easy and difficult transportation have an impact on the value of the house.)
8) Heating type (The method used to heat the building affects the price.)
9) Facades (The facade feature of the building affects the price.)
10) Green area (This feature positively affects the value of the house.)
11) Seascape (Seafront residences are always more expensive.)
12) Landscape (This feature positively affects the value of housing.)
13) Security system (The security system in the building increases the value of the house.)

14) Balcony (One of the key features sought in the housing. The width of the house and the number of balconies being more than one may positively affect the value of the house.)
15) The Neighborhood (The House's neighborhood is very important in terms of value. In good and distinguished neighborhoods, the price of a house is always higher.)
16) Indoor and outdoor parking lots (The fact that the house has a parking garage or an outdoor parking area has a positive effect on the house's value.)
17) Number of floors and the house's floor (It can affect the value of the house positively or negatively.)
18) The Building's age (The Building's age is effective in the value of the house. Usually, as the building's age increases, the value of the house decreases.)
19) Smart home systems (These systems positively affect the value of the house.)
20) Sports facilities (Positive effects.)

1.2 Turkey´s Real Estate Market

House prices in Turkey decreased by 2% between 2007 and 2011. This was mainly caused by the economic growth slowing to 0,7% in 2008. Housing prices decreased by 14,65% after the inflation in 2008, by 2,82% in 2009, 3,54% in 2010 and 2,39% in 2011. Since then, the real estate market has grown steadily and strongly in terms of value. Despite some deceleration, this was expected to continue throughout the forecast period. The growth in Turkish real estate has in part been driven by extensive urban renewal and development projects. Adding to this is the increase in foreign investment in the market, aided by the abolition in 2012 of reciprocity laws that inhibited investors from certain countries buying land and property in Turkey. The Turkish real estate market grew by 5.1% in 2014 to arrive at a bulk of 9.261 thousand units. In the 2010 -14 term, the ratio of its annual combined increase was 3.6%. Turkey constitutes 4.9% of the European real estate market value. Turkey´s real estate market continues to perform strongly, despite price rises slowing since the beginning of 2016 (Global Property Guide, 2016).

As reported by the Central Bank of the Republic of Turkey (CBRT), the house price index rose by 13.98% throughout July 2016 countrywide. In Istanbul, Turkey's largest city, the house costs dramatically increased with the level of 17.68% in 2016. In Ankara, the capital city of Turkey, house prices showed some fluctuation with 9.16% in 2016. Besides, there was an increase of 4.71% in real estate costs in Turkey's third-largest city; Izmir.

Moreover, according to information from CBRT, the highest annual rise in newly built housing prices has been seen in Istanbul by 15.33% during the year to July 2016, followed respectively by Izmir at 11.91% and Ankara at 11.55%. In 2015, dwelling sales expanded in Turkey by 10.6% (1.3 million units) but it decreased by 2% in the first eight months of 2016, based on the information given by the Turkish Statistical Institute (TSI). Coming to foreign sales, the Turkish government made the foreign property's rules easier in 2012, and the foreign buyers became more important. The sum of sales to foreign investors that declined by 17.3% before then, increased by 20.4% (22.830 units) in 2015. Foreigners bought houses mainly in Istanbul with 7.493 with purchases, Antalya (6.072 purchases) and Bursa (1.501 purchases).

The real estate market of Turkey, with a rise of 22% is estimated since 2014 to be worth $ 37.2 billion. The ratio of annual combined increase in the market is expected to be 4.1%, in 2014-19 (Invest in Turkey, 2016). Hence, real estates are seen as a good option for secure investments, particularly for ordinary people. Besides, real estate can be called an investment that does not fall rapidly. But, the recent stagnation represents that its prices cannot always continue to increase.

1.3 Housing Market

The housing, which is one of the indispensable needs of all societies since the establishment of the world, can be defined as a long lasting physical space that meets the housing needs of people. Besides providing shelter, the house has a social quality in terms of its features such as a place where the socialization phase of individuals and family relations pass, privacy, belonging and its sense of residence. In other words, the house has many functions such as being a shelter, being an investment vehicle, contributing to economic development, being an assurance for the future of individuals, and it can also be considered as an indicator of economic and social development of societies. It also has an important place in affecting the entire economy by its implication to monetary factors.

It is also possible to define the housing sector as a locomotive sector. Because the sector is based on a significant amount of domestic capital, the employment potential is also high. However, its superiority in terms of creating value-added and the fact that it is in an input and output relationship with other sectors, especially manufacturing, increases the

importance of the sector. Besides, an increase in housing expenses due to the high multiplier effect of housing expenditures leads to an increase in the demand for home-related goods such as white goods, furniture, and home textiles (Öztürk, 2009).

The housing market, which has an important place in the economy, is also a place where housing services are allocated through a supply and demand mechanism. As the housing is both a property and an investment vehicle, the housing market has some differences compared to other markets. These differences can be listed as being very costly housing supply, permanent residence, being heterogeneous, being static, causing growth in secondary markets and being used as a guarantee. Also, unlike other goods industry, the housing market is unique, because it highlights structure, location, and environmental characteristics. In other words, a house's features affect its value positively or negatively. For example; house size, number of rooms, parking, transportation, heating, etc... These are important characteristics in determining the value of a house. Also, employment, population growth, building costs, and real interest rates affect housing prices.

The price of a house emerges as "Rent and purchase price". For this reason, the share that households can allocate for housing for their rent or purchase is forming the basis of the demand for housing. The housing demand of households which are considered to have sufficient economic power, just as the demand for other goods and services, is affected by the income, prices, expectations, tastes, preferences, complementary and substitute goods. In other words, housing preference varies according to the socio-cultural expectations of individuals and demographic structures. The housing is not entirely considered by the consumers as a material asset. On the axis of this theory, it is possible to talk about some important variables that are effective in the selection of housing for individuals.

These are (Göncü, 2004: 133-134):

i. First of all, the location is very important; which area, which neighborhood or which street,
ii. What is the situation of taking the sun,
iii. What is the rental income,
iv. Whether they have environmental regulations such as green areas, indoor and outdoor car parks, children's playgrounds, sports grounds, shopping centers,

v. The appropriateness of the housing for a social purpose, the predictions of the property environment,
vi. Safety and accessibility in terms of roads, water, electricity, and infrastructure,
vii. Traffic density, air pollution and volume of sound level,
viii. The characteristics of the neighborhood (neighborhood relations, whether the neighborhood is elite, occupational groups, etc.),
ix. Distance to public spaces such as city center, hospital, school,
x. What are the settlement criteria and total usage areas,
xi. Whether there are security systems (caretaker, security guard, security cameras).

These are important factors in housing demand. These factors and changes in housing preferences (such as duplex, triplex, closed sites, studio type apartments, etc.) increase or decrease the housing prices. The change of real estate prices that affects socioeconomic conditions is a source of worry for both persons and governments and has an additional impact on national income conditions. The expectations of capital profit from investments in housing will influence housing prices with the increase in housing requests and will lead to elevated fluctuations in house prices. As a result, if the housing demand is not corrected in the short term, it will lead to a rise in property prices. The amount of material used in the house, the quality and the size of the house, its location and infrastructure, the labor it required, the building construction cost, and land price are other factors affecting housing prices. The heterogeneity of housing prices can be attributed to the fact that the factors listed are different.

1.4 Research Objectives and Contributions

The real estate is very complex due to its structure and it contains many features and it is caused that there are difficulties in obtaining realistic estimations. Traditional methods may be incomplete in this field and show less performance than non-traditional methods. One of the aims of this study is to obtain more accurate price estimations by using non-traditional methods and to examine the characteristics that affect the real estate price. In addition, it is intended to develop a unique hybrid approach for price estimation using the SelectKBest, Grid Search and artificial neural network (ANN) methods which are the non-traditional methods and then apply them to districts' data sets of Istanbul. Selecting of most important housing features that affect the price with SelectKBest method, optimizing of the appropriate hyperparameters for ANN with Grid Search method, obtaining of the best estimation

performance and the smallest error values using ANN method are among the important objectives of the study. In literature studies, in general, linear regression method and the ANN methods were compared and it was stated that the ANN method was most of the time more successful. Nghiep et al. (2001) have made comparisons between the ANN model and MRA models. They argued that the ANN model would obtain more successful results if sufficient data set and the correct data were selected. Otherwise, they stated that it might not be probable to reach a clear conclusion about the success of the ANN results. This case emphasizes the importance of the size of the data set and accurate data preprocessing for ANN models. In this thesis, the preprocessing steps made while preparing the data set are considered to contribute to the performance of the model. For the best possible results, the integrity and accuracy of the data were taken into account. Zurada et al. (2006) stated that when there is enough data and accurate analysis, fuzzy logic and ANN models are useful models and should be studied. Again, the importance of the data set and the size of the data set for this method is emphasized. The large number of data sets studied in Istanbul provides an advantage for the results of the study.

In research, the studies with ANN were generally successful. Compared to linear models, better results were obtained. Selim (2009), made one of the research on real estate prices estimation in Turkey. In his study, using the data sets created by the survey results in Turkey, he found out that the ANN model's ability to predict was successful and stated that it would be a better model for the prediction of real estate price in Turkey.

It is clear that with more work and development in this area, better results will be obtained. Considering the ability of heuristic methods on complex data, it is predicted that they will achieve success with accurate analysis for housing price estimation. The success of the estimation results obtained from the intuitive method developed in this study will be an important reference for future studies.

1.5 Overview

The other parts of this thesis continue as follows: The literature studies are given in the next section. The literature section is divided into 3 sections. The studies on real estate valuation are classified according to methods. In section 3, detailed pieces of information about k-fold cross-validation, SelectKBest selection method, Grid Search, ANN methods are

available. In section 4, there are the steps and analysis results of the SelectKBest, Grid Search and ANN methods applied in this study. Finally, section 5 provides explanations of the results obtained with this study and the contributions and suggestions for further studies.

2. LITERATURE REVIEW

Recently, there are many studies on the evaluation of housing prices. Various methods are proposed to forecast the value of the real estate market. Pagourtzi et al. (2003) classified the methods as traditional methods and advanced methods to identify the house's purchase price or rentals. He states that there are methods such as multiple regression analysis method, investment/income methods, regression method in the traditional ones and he adds that there are methods such as ANN, hedonic price method, fuzzy logic in the advanced methods.

It is also possible to classify advanced methods as non-traditional methods. In the field of real estate price valuation, the use of non-traditional methods is increasing. In this section, the use of non-traditional methods in the literature for housing price estimation is examined and these studies are classified as follows; section 2.1 presents studies with the Hedonic Methods, Section 2.2 presents studies with Adaptive Network-Based Fuzzy Inference (ANFIS), Fuzzy Logic and ANN methods, Section 2.3 presents studies with Genetic Algorithm.

2.1 Researches of Hedonic Model on Real Estate Price Estimation

The fact that housing has a heterogeneous feature has increased the use of hedonic methods in price estimation studies and it can be said that it has an important place in research and caused many studies in this subject (Limsombunchai, 2004; Peterson, Flanagan, 2009; Sayer, Moohan, 2007; Peterson, Flanagan, 2009; Jiang, Phillips, Yu, 2014). Limsombunchai (2004), using data from 200 homes collected from a website in New Zealand for the data set, used non-traditional models in the real estate price estimation study. These models are hedonic models and ANN models and compared the data from the models and analyzed the results. According to the results of his analysis, he indicated that the model of ANN is better for real estate price estimation, and this determination was made according to the performance criteria of the highest R^2 and the lowest root mean square error (RMSE). Besides, he stated that hedonic price models give poor results in non-sampling sections and that experimental indicators support the power of neural networks in housing price estimation. Sayer et al. (2007) conducted a study using the hedonic method for housing price estimation. The data set of his work was based on data from East Midlands and the United Kingdom. They also used

regression analysis in their studies. Their regression analysis studies were divided into three models and analyzed the regression analysis results of these three models statistically. They stated that the R-squared value was higher than the other suitable linear models and which the reason for this could be a working space environment and width. In his work, he stated that the forecasting ability of the hedonic method is as low as described in one of the three models they allocate. As in the study of Limsombunchai (2004), Peterson et al. (2009) compared the ANN model with linear hedonic model and analysis results. They used approximately 46.000 housing data in their research. They stated that linear hedonic pricing models produced statistically greater pricing errors and the magnitude of this error increased over time. Teixeira et al. (2010) conducted a house price forecast study with the hedonic method, using a data set consisting of data from a Portuguese city. These data consisted of new and non-new houses and included various internal and external qualities of houses. In their research, they emphasized the classification of the properties of the houses and the effects of these properties on the house price. Jiang et al. (2014) developed a new hedonic model for price forecasting between 1995 and 2014. They compared the extrapolated success of the model with the indices using the Case and Shiller method (1987, 1988). As a result of their study, they specified that their method performed better than the S&P/Case-Shiller index at estimating the single sale house price. At the same time they indicated that it showed poor performance in forecasting the price of repeat-sales house. Yayar et al (2014) examined the factors affecting the prices of real estate in Turkey and the hedonic model of residential properties on the value of house has studied the effects of positive and negative. Three different models were used in their studies and they separated these models as follows; linear, semi-logarithmic and full-logarithmic. They stated that according to the results of his models, the characteristics of houses could increase or decrease the value of the house.

2.2 Metaheuristic Approaches to Real Estate Appraisal (ANN, Fuzzy Logic and ANFIS Model)

ANN model has been studied extensively in the literature for home price evaluation and some of these studies have been compared with multiple regression analysis (MRA) methods. ANN has been used in the estimation of housing prices since the early 1990s. The first study by Borst (1991) was followed by other studies (Worzala et al. (1995); Do and Grudnitski, 1992; McGreal et al. (1998); Bee-Hua, 2000; Tay et al.1992; Lenk et al. (1997)). Tay and Ho (1992) examined the predictive power of apartments with the ANN. Tay and Ho

(1992), have conducted a study using the ANN model for apartment price forecasting. They used a data set with more than 1.000 sales data. They compared the results of the analysis using traditional MRA and ANN model. According to the results of this analysis, they stated that the mean absolute error (MAE) of the ANN model is lower than the MAE of the regression model and the neural network model could provide a better estimation than multiple regression analysis. Do et al. (1992) have made observations similar to the results of Tay and Ho (1992)'s research. Do et al. (1992) used 105 residential data and stated that the ANN model was good at predicting the real estate value. They compared pricing errors of ANN models and linear models. By comparing the results, they found that the ANN model pricing errors were smaller. Rossini (1997), using the ANN model in his research on housing valuation used three general procedures. He compared these procedures with the results of the MRA model. Parameters used in the data set; sale time, sale price, neighborhood, zone, improvement, land area, number of rooms, wall type, roof type, status, equivalent construction area, building type and building construction history. As a result of the analysis, multiple regression analysis reached 90% average fit and 89% accuracy, and ANN methods remained at 78% mean fit and 81% accuracy. Do et al. (1992) and Tay et al. (1991) showed different results against the findings of both Worzala et al. (1995) and Lenk et al. (1997). Worzala et al. (1995) investigated the performance of the ANN model using more than 250 data in Fort Collins and Colorado. They examined the results obtained from the MRA model and ANN models. According to the results of their analysis, although the ANN models were slightly better than the MRA models in some cases, the results of the artificial neural network model and the MRA model in some cases were not different. Lenk et al. (1997) used approximately 280 house data as data sets. They compared the hedonic method and ANN models in their study and according to the results of the analysis, they argued that both models had similar performance in contrast to the recent research on the superiority of ANN models. McGreal et al. (1998) examined the ANN model in real estate valuation. Using data on market sales, they evaluated the ability of the model to predict in a test sample. They were skeptical about the potential of the neural network model and they also took into account Worzala et al. (1995) research on real estate forecasting. Bee-Hua (2000) used a hybrid model by combining an ANN model with a genetic algorithm (GA) model to estimate residential construction demand in Singapore. He said that the ANN model is safe to use and can combine two models to produce better and more accurate models. For Bee-Hua (2000), the results have displayed an exceptional development in estimating the correctness based on the

decrease of the average mean absolute percentage error (MAPE) with this united method and the benefits provided by the techniques of these methods (such as optimization and search techniques). Also, he indicated that the duration of completion of the united method training was long and stated that this was a disadvantage for GA. KAUKO et al. (2002) conducted a study using the ANN on the real estate market in Helsinki. As a result of his studies, he stated that the ANN technique demonstrates the skills of classification. Ge et al. (2003) utilized the ANN model to estimate the prices of specific housing properties using data between 1980 and 2001 in Hong Kong. According to the results obtained, they stated that the ANN method has a good predictive power and also has a superiority in mapping the complex nonlinear relationship between variables. Khalafallah (2008) analyzed the sales of real estate by ANN methods. The parameters used are; time, interest rate, change of sales unit value over the previous year, sales by years, average sales time and transaction volume. According to the model results, it was observed that the estimation errors were between -2% and +2%. Lam et al. (2008) investigated the performance value of the integration of Entropy and ANN models for housing price estimates. They used 4.143 data sets from the Centaline Property Agency Limited in Hong Kong website in their studies. According to the results of the analysis, it is stated that the model has a better function for estimating the real estate price in the case where this model has more appropriate parameters and relatively small sample size. Selim (2009) also argues that the ANN model is convenient. He used the data in 2004 which has formed on top of the 5.000 survey results in Turkey and he analyzed the ability to predict the hedonic regression models and ANN methods in real estate evaluation. According to his study's results, compared with the hedonic regression to estimate the price of housing in Turkey it has stated that the ANN methods may be a better alternative. Ocerin et al. (2013) conducted a study of ANN and a hedonic model for estimation of real estate in Spain. As a result of the analysis, they predicted that the estimates are well adapted to the real market conditions, and may also be estimated in extreme cases. Khamis et al. (2014) applied two methods (ANN and Multiple Linear Regression (MLR)) to estimate property prices using data collected from a website in New York and compared the performance of these models. Khamis et al. (2014) approved the approach of Do and Grudnitski (1992) and concluded that the ANN model could be preferred as an alternative model for estimating property prices compared to the MLR model.

The fuzzy logic has been widely used in many engineering studies and it can be said that there are many studies performed with this model for real estate price estimation. Also, studies with fuzzy logic have achieved successful levels. (Kuşan, Aytekin and Özdemir, 2010; Ustundag, Cevikcan, Kilinc, 2011; Del Giudice, De Paola, Cantisani, 2017; Mukhlishin, Saputra, Wibowo, 2017; Byrne, 1995). Byrne (1995), using the fuzzy logic model, examined the utility of this model in real estate evaluation. He also stated that the model is a good way to overcome uncertainty in the evaluation of housing prices. Bagnoli et al. (1998) examined the feasibility of the fuzzy logic model in the evaluation of housing prices. They argued when the fuzzy system and linear regression were compared, the estimated sales price produced by the fuzzy system was more accurate. Kuşan et al. (2010) researched the fuzzy logic model by using 200 data from different regions of Eskişehir to forecast the housing price. They compared the actual prices with the forecast values and stated that the model results were successful. Ustundag et al. (2011) used the hybrid method of Fuzzy Rule-Based System (FRBS) which is integrated with the Fuzzy Analytic Hierarchy Process (FAHP) for housing price estimation in Istanbul. Firstly, they tried to estimate the meter square (p) of three different houses in different districts of Istanbul and then compared the estimated market prices of the houses with the same characteristics to the average market prices of the houses. They stated that the percentage deviation rate ranged from 3% to 8%, and this model could be noted as a useful forecasting model for the housing price. Mukhlishin et al. (2017) have prepared a study using K-Nearest Neighbor, Fuzzy Logic and ANN methods to estimate house prices in Indonesia. They looked at the results of these methods to find the most appropriate price, they said that there was a significant difference between the methods and that the fuzzy method was better in estimation accuracy.

The ANFIS model, which uses the learning ability of ANN methods and uses the deduction of fuzzy logic, has some studies in the field of real estate appraisal (Hasiloğlu et al., 2004; Guan, Zurada, Levitan, 2008; Gerek, 2014; Sarip, Hafez, Daud, 2016). Guan et al. (2008) performed a study using the ANFIS model for real estate price estimation. They compared the results of the ANFIS model with the results of the traditional regression approach and stated that ANFIS was a viable and useful method for estimating real estate price. Gerek (2004) has prepared a study to estimate real estate prices using two ANFIS models. These models are as follows; the ANFIS-GP model, the ANFIS-SC model. He used the data of 91 housing prices in Turkey in his research. He argued that the grid partition

technique could give better results in evaluating the mean MAE values. Furthermore, Gerek (2014) pointed out the effects of some factors on the accuracy and predictive power of the model and reported the positive and negative effects of these factors on real estate prices. Sarip et al. (2016) conducted a study for estimating the prices of houses. They used fuzzy least-squares regression-based (FLSR) method in their study and they also applied ANFIS and ANN models in their studies and then compared the performance of these models. They stated that the highest decrease in MAE in FLSR was better than ANFIS and ANN according to the test results. As a result, they indicated that fuzzy estimations produced by the regression-based model are more accurate than ANFIS and ANN models for estimating the house price.

2.3 Genetic Algorithm Studies for Real Estate Forecasting

Although the Genetic Algorithm is a stochastic technique used in many areas, it has not been studied extensively in real estate. In some studies in which GA model is used in the literature, it has been analyzed by making comparisons with other models or by creating a hybrid model with another model. Ng et al. (2008) have been compared results of four models using data 5 to 10 years to make home price forecasting. These models; GA, Genetic Algorithm with Linear Regression Analysis (GA-LRA) model, the Linear Regression Analysis (LRA) model, the GA-LRA model with the Adaptive Mutation Rate (AMR). According to the results of their studies, they observed that the LRA model could not make successful predictions. Besides, they noted that the genetic algorithm method has a lower estimation performance in the 5-year data than the 10-year data and the fourth model (the GA-LRA model with the AMR) has the best estimates. Gu et al. (2011) attempted to estimate the average sales price by using house data in China. They used a hybrid model (advocated vector machines and GA) for estimation and compared the real estate price estimation performance of this model with results of gray model (GM). As a result, they indicated that the genetic algorithm method spent less time and the predictive accuracy of the genetic algorithm and support vector machines (G-SVM) method was better than GM method. Ahn et al. (2012) examined to estimate home sales index (HSI) and home rental index (HRI) in the Korean housing market using ridge regression coupled with genetic algorithm (GA-Ridge) method. They used data between 1996 and 2009 from a bank in Korea for their study. They compared the results of the GA-Ridge with the results of MLR, Pure Ridge regression, and ANN. According to the analysis the results of HSI, they stated that the performance of the methods was remarkably different and that the GA-Ridge method had better performance than other

methods. Manganelli et al. (2015) examined the relationship between the geographical location and price of properties, using the GA model. Their data set included 190 houses in the city of Potenza. They compared the results of the GA model with the results of the MRA model. As a result of the comparison, they proposed that the marginal prices were different from slightly each other and nevertheless both models produced statistically similar results. At the same time, they mentioned that the GA model has a good skill to interpret statistical information (the error information). Del Giudice et al. (2017) evaluated real estate rental prices with a GA model by using data from a Naples region. They examined the results of the GA and MRA models to demonstrate the predictive potential of GA and to determine the reliability of GA in real estate market analysis. They noted there is no significant difference between them according to the models' analyses results besides the GA model has the superiority of interpreting the housing rental values. Del Giudice et al. (2017) and Manganelli et al. (2015) pointed out similar results in their study.

3. MATERIAL AND METHODS

3.1 Method

In this study, in order to obtain more accurate estimation values and to facilitate the selection of properties affecting the estimated value, feature selection method (SelectKBest) was used. A hybrid system has been developed by using ANN and feature selection method to estimate real estate sales prices. In addition, the grid search method was used to optimize the ANN model hyperparameters. Also, k-fold cross validation technique is one of the methods in this study. After giving preliminary information about the correlation analysis, detailed information about these methods used in this study is presented under related headings.

3.2 Correlation Analysis

Correlation is a statistical relationship used to determine the linear degree of the relationship between two independent variables. If the value of two or more variables (e.g. X, Y…) varies according to each other's value, it can be said that there is a relationship between these variables. For example, as the age of a house increases, the price may decrease. However, the fact that the house is a historical monument can make it more valuable. An ancient or antique monument in Istanbul gains value year by year. The location and the direction that the house faces are also key factors to the housing price. For example south-facing houses are more valuable in Adana (fifth biggest city in Turkey). This may not always be meaningful or accurate. The residence's surface is another important factor to determine the value of a house due to the fact that there is a strong and positive correlation coefficient between the two (the house and the surface).

Correlation measures the strength of the linear relationship between variables. For example, in a scatter diagram, the proximity of the points to a straight line increases the strength of the linear relationship between the variables (Bewick at al., 2003). The degree of

this linear relationship is determined by the correlation coefficient. The degree of this linear relationship is determined by the correlation coefficient. The coefficient of correlation is indicated by the letter "r", which takes a value between (-1) and (+1). The sign of numbers indicates the direction of the relation between variables (positive or negative), and the magnitude of the numbers indicates strength of the relation. If the correlation is positive (+), the variables have changed in the same direction. A correlation coefficient of (+) indicates that the two variables are in a relationship in the same direction, and a negative (-) indicates an inverse relationship between the two variables. Correlation analysis is one of the methods used to determine the aspect and force of the relation between variables, regardless of whether the variables are dependent or independent. The correlation coefficient takes the value (0), if there is no relation between the variables, and it takes (+1), if there is a positive and complete relation, and it takes (-1), if there is a negative and complete relation.

Correlation coefficient can be calculated with different formulas. One of the formulas is shown below;

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{[\sum(x-\bar{x})^2][\sum(y-\bar{y})^2]}} \quad (1)$$

3.3 Dimension Reduction Method

Dimension reduction process is a meaningful expression of higher dimensional data in a lower dimensional space. The dimension reduction methods can also be used for feature selection. The main aim in dimension reduction is to detect the unnecessary properties of the available data and discard it from the data. Dimension reduced data is easier to process. The dimension reduction process can be partitioned into attribute extraction and feature selection sections.

3.3.1 Feature Extraction

For any problem to be solved by machine learning methods, the system must be appropriately represented. The characteristics of the problem to be solved may not always be appropriate for machine learning methods. The purpose of feature extraction is to get the best information from the original data and represent it in a lower dimensionality space. Feature extraction (characterization) operation is a process of dimension reduction. Accordingly, the

complexity of a complex data is reduced to a simpler problem. The true feature extraction and its appropriate system design are the factors that affect the success and performance of the result.

3.3.2 Feature Selection

Feature selection is the process by which the best features are selected to determine which features within the data set are effective in the results. Feature selection that is also called qualification or variable selection is the operation of selecting the best features between the features in the data set by appraising the properties according to applied the method (Forman et al., 2003). In the selection of features, it is significant to reduce the number of features in the data set by selecting the most effective and most important features according to the desired result. Reducing the number of features provides many advantages to practitioners during the analysis process.

Advantages of the feature selection process (Ladha et al, 2011);

1. Decreases the size of the feature set and facilitate analysis
2. Eliminates non-relevant and noisy data,
3. Improves data quality,
4. Makes the data set simpler, visualized and understandable,
5. Saves the resources during the data collection process,
6. Reduces the amount of memory required to store data,
7. Increases the success of the obtained model.

The methods used in feature selection are generally grouped into three groups. These are filtering methods based on statistical information, wrapper methods that perform search operations on features, and embedded methods based on finding the best dividing criterion (Saeys et al, 2007).

3.3.2.1 Filtering methods

Filtering methods are known as the oldest feature selection methods used in data mining. In these methods, feature selection is made by using functions based on statistical

criteria such as distance, information, dependency and consistency measurements without using any classifier. In these methods, which work with similar logic, a value (score) is calculated for each property in the data set by the evaluation function and the properties with the highest values are selected to the best subset of these values.

The Scikit-Learn library, which is a Python program library, is used frequently in machine learning and process such as feature selection, size reduction can be done with the methods in this library (Pedregosa et al, 2011). The "SelectKBest" method, which is used as a filtering method from the feature selection methods of the Scikit-Learn library, aims to determine the best properties associated with the target property by using statistical tests to select a certain number of features. The scores of the features obtained from these statistical tests are found and features with the best scores are selected according to the specified number. Depending on the type of problem, score functions such as chi2, f_regression which are the functions of "SelectKBest" method are used. Generally, f_classif and chi2 are preferred for classification problems and f_regression is preferred for regression problems.

1) "f_classif" function uses the ANOVA f test and outputs a score known as the f value. The ANOVA test gives the ratio of variance explained by the features.
2) The function known as "f_regression" is used for regression type problems.
3) "chi2" (chi-square) function produces scores according to the chi-square test based on whether the difference between observed and expected frequencies is significant.

3.4 Hyperparameter Optimization

It is a common problem in machine learning that parameter values are needed to initiate each supervised or unsupervised learning algorithm to be used before training. These parameters are called hyperparameters and can cause very effective changes to the model. They are used to configure many outcomes such as learning rate, kernel parameters, network architecture, etc., for a given learning algorithm. Before the model is created, the optimization of hyperparameters positively affects the success of the model.

Hyperparameters that may vary by the model should be determined before the learning process begins. For example, parameters such as support vectors for a support vector machine, coefficients for linear regression or logistic regression, activation functions for ANN methods are important model hyperparameters. Many models have important parameters that

cannot be estimated directly from the data. For example, as in the k-nearest neighbors classification model. This type of parameter is called a fitting parameter because there is no analytical formula for calculating an appropriate value (Kuhn et al., 2013).

3.4.1 Algorithms for Hyperparameter Optimization

When designing a model, the initial selections for hyperparameters may not usually lead to accurate results. Iteratively, the success of the model is observed by changing the hyperparameters one after the other and the most suitable hyperparameter group is selected for the model.

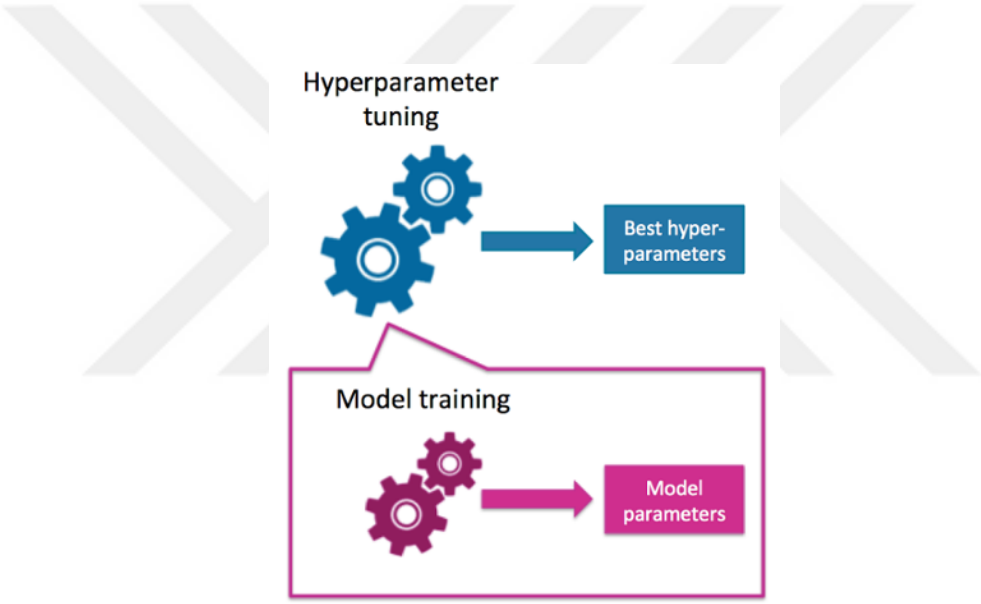


Figure 3.1. Hyperparameter tuning

In the intuitive parameter selection, they are estimated by using preliminary information about the problem, the model is designed according to these hyperparameters and the results are observed. Intuitively, the model reconstructions by making new hyperparameter estimates that will increase the performance of the model according to the results, then it is trained and the results are observed. A limited number of parameter attempts can be made in heuristic search, and success is limited to predictions about the problem. In addition to the intuitive selection of hyperparameters from experience with previous problems, it can be used in different techniques for optimal selection of the most appropriate

hyperparameter group. Grid Search, Random Search, GA and Bayesian methods are among the methods used for hyperparameter optimization.

3.4.2.1 Hyperparameter Optimization with Grid Search

Some of the hyperparameters are able to take an infinite number of values. However, by using preliminary information about the problem, ranges can be determined for the values that hyperparameters can take. By selecting specific main points from these specified ranges, value lists for hyperparameters are created. When selecting hyperparameters with Grid Search; the network is trained for all combinations of values within the specified range and the results are observed, the best combination is selected as the hyperparameter group, depending on the situation. Bergstra and Bengio (2012) say that the grid search method will give researchers insight and is easy to implement.

Considering that a machine learning model takes the hyperparameters A, b1, b2 and b3, in the grid search, the range of values for each of the hyperparameters a1, a2 and a3 is first defined. This can be considered as a set of values for each of the hyperparameters. The grid search technique will form several versions of A with all possible combinations of hyperparameter values (b1, b2 and b3) defined first. This range of hyperparameter values is called the grid.

Suppose the grid is defined as follows;
b1 = [15,20,30]

b2 = [1,3,5,7,10,90]

b3 = [100,200,300,400,500,600]

The grid search will begin the process of creating several versions of A with the previously defined grid. And this search will start with a combination of [15, 1, 100] and end with [30, 90, 600]. All searches between these two will go through combinations, this may result in a long computational process for the grid search. In addition, parallel operations can be carried out in order to test different combinations of parameters in the grid search, thus saving time can be ensured. It can be studied on a subset of data set for parameter selection and thus again saving time can be ensured. The purpose of this process is not to find the most

appropriate values for hyperparameters, but to create a general opinion. Thus, it determines which ranges of hyperparameters to focus on. When determining hyperparameters, it is possible to initially keep the parameter values large and then reduce them to the appropriate range of values, it achieved earnings from time thanks to action taken in this way. For example, selecting a range such as [10, 50, 100, 500] for the number of neurons in the artificial neural network initially, and then expanding the scope according to the results, specifying a range such as [200, 250, 300, 350, 400] will save time.

3.5 Artificial Neural Network Model (ANN)

ANN models, one of the applications of artificial intelligence, is an information processing system that analyzes existing data by mimicking the working structure of the human brain (Fausett et al, 1994). The human brain, working principle, and properties have been researched for many years. ANN model has emerged as a result of efforts to artificially model the working system of the human brain. The first ANN model was developed in 1943 by Warren McCulloch and Walter Pitts (McCulloch et al, 1943). Then, B.G. Farley and W.A. Clark have developed a model that responds to warnings within a network and can adapt to warnings, in 1954. In 1959, Bernand Widrow and Marcian Hoff developed the artificial neural network models named ADALINE (Adaptive Linear Neuron) and MADALINE at Stanford University (Widrow et al, 1960). The ADALINE model is the basis for the subsequent neural network studies and has the same characteristics as the Rosentblatt perceptron model. Also, it has a more advanced learning algorithm model. MADALINE has been used as a filter to eliminate the sound echoes in the telephone lines and it is the first network that has been applied to the real problems still. The first neural computer appeared in 1960. These models are the first studies of ANN in engineering applications. In 1963, the first deficiencies of the simple models appeared and then the beneficial results were deferred until the theoretical structures of thermodynamics in the 1970s and 1980s were used in the development of nonlinear networks. ANN models began to gain more recognition in 1985, and intensive research on this subject began (Mehra et al, 1992).

In 1988, Broomhead and Lowe developed the radial-based function model. This model that is developed as an alternative to the multi-layer perceptron model was particularly successful in filtering problems. The process of the development of ANN models still continues.

ANN models are often used to perform the following functions in applications; Estimation: The output values are estimated by using the examples introduced in the ANN for the input values. Data Filtering: The networks trained for this purpose fulfill the task of identifying the appropriate ones among many data. Failure Determination and Diagnosis: The networks developed for this purpose are used in detecting the problems occurring in processes in the machines, the systems or in the most general processes and in determining the faults. Classification: ANN models are used successfully in classification applications. In these applications, the samples are clustered into specific classes, and then a sample that belongs to which class is determined. Completion of Missing Data: The networks trained for this purpose determine whether the data is inaccurate and incomplete. It completes the missing information.

ANN models are inspired by the human brain and come to exist from the mathematical modeling of the learning process. ANN, which imitates the human brain works simplistically has many important features such as learning from data, generalize, work with an unlimited number of variables and so on. In Figure 4.1, the structure of a nerve cell (neuron) is given**.** For this reason, studies on the ANN models started with the modeling of neurons, which are the biological units that form the brain, and their applications in computer systems. Later on, ANN models are used in different fields in parallel with the development of computer systems.

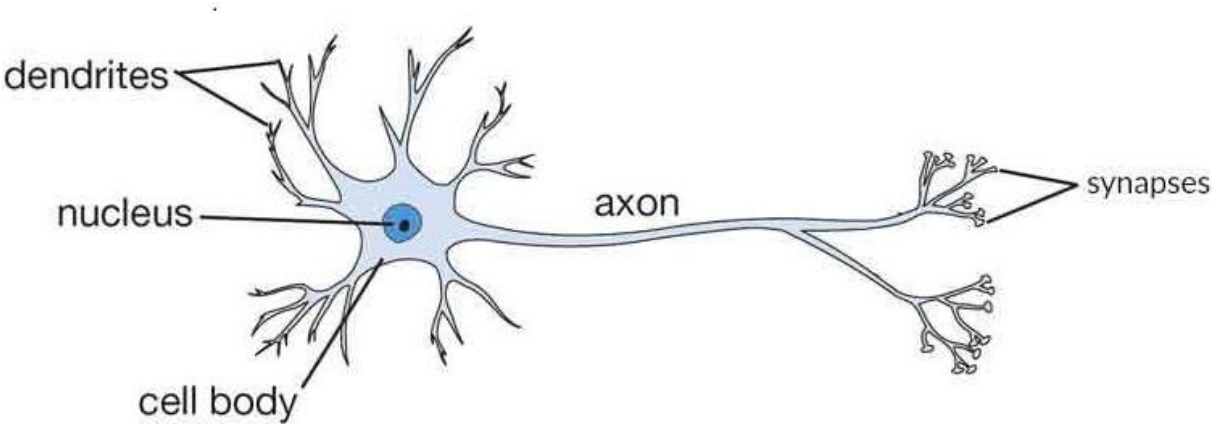


Figure 3.2. Biological Neuron

3.5.1 Structure of Artificial Neural Networks

The main unit of ANN models is nerve cells. In engineering disciplines, artificial nerve cells are also named processing elements. Each element has five main elements. These are inputs, weights, activation function, aggregation function and output. As shown in the figure, each cell has a multi-input single output structure. (Tsoukalas et al, 1996)

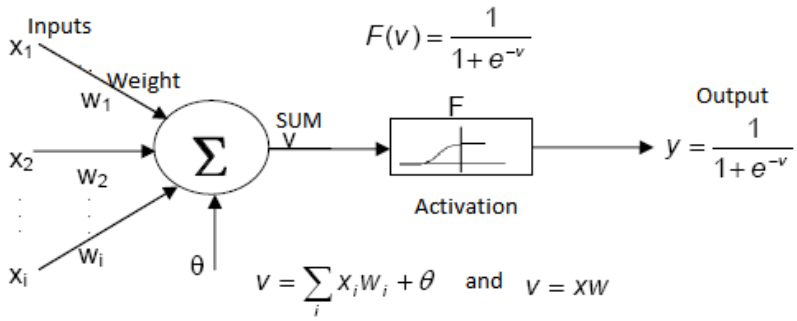


Figure 3.3. Artificial neural cell (artificial neuron)

1. Inputs (x_1, x_2, ..., x_n): A set of data created by the user with instances for cells in the input layer. For cells in the other layer, it can be the output of the cell in any layer.

2. Weights (w_1, w_2, ..., w_n): Show the amount of the transfer from inputs to the output. For example, the weight w1 indicates the effect of the input x1 on the output. The fact that the weights are large, small, positive or negative does not indicate that the relevant entry is significant or unimportant. The weights may be fixed or variable values.

3. Aggregation function: Used to calculate the net input of a cell. Different functions are used for this purpose. The most preferred is the weighted total function. In this function, each input is multiplied by its own weight and these values are gathered. The aggregation functions of all cells in the ANN do not need to be the same. Each cell can have a different aggregation function independently. The other aggregation functions that are used are shown in Figure.

Sum $\text{Net} = \sum_{i=1}^{N} X_i * W_i$	*The weight values are multiplied by the inputs and the calculated values are added to each other to calculate the Net input.*
Max $\text{Net} = \text{Max}(X_i * W_i)$	*Within n inputs, the weights are multiplied by the inputs, and the largest of them is considered to be the Net input.*
Min $\text{Net} = \text{Min}(X_i * W_i)$	*Within n inputs, the weights are multiplied by the inputs, and the smallest of them is accepted as Net input.*
Weighted total $\text{Net} = \text{Net}_{(previous)} + \sum_{i=1}^{N} X_i * W_i$	*In this function, each input is multiplied by its own weight and these values are summed.*

Figure 3.4. Aggregation functions used in the artificial neuron cell

4. Activation Function: Used to calculate the output value to be generated in response to the net input value of the cell. As in multilayer perceptron, the activation function must be differentiable in some neural network models. In addition, the calculation of the activation function's derivative is important in terms of the duration of the training of network. The derivative of the sigmoid function can be written in terms of the function itself. The sigmoid function is widely used because of this providing easy operation. As with the aggregation function, not all cells need to use the same activation function. Each cell can have a different activation function independently. The following figure shows the various activation functions.

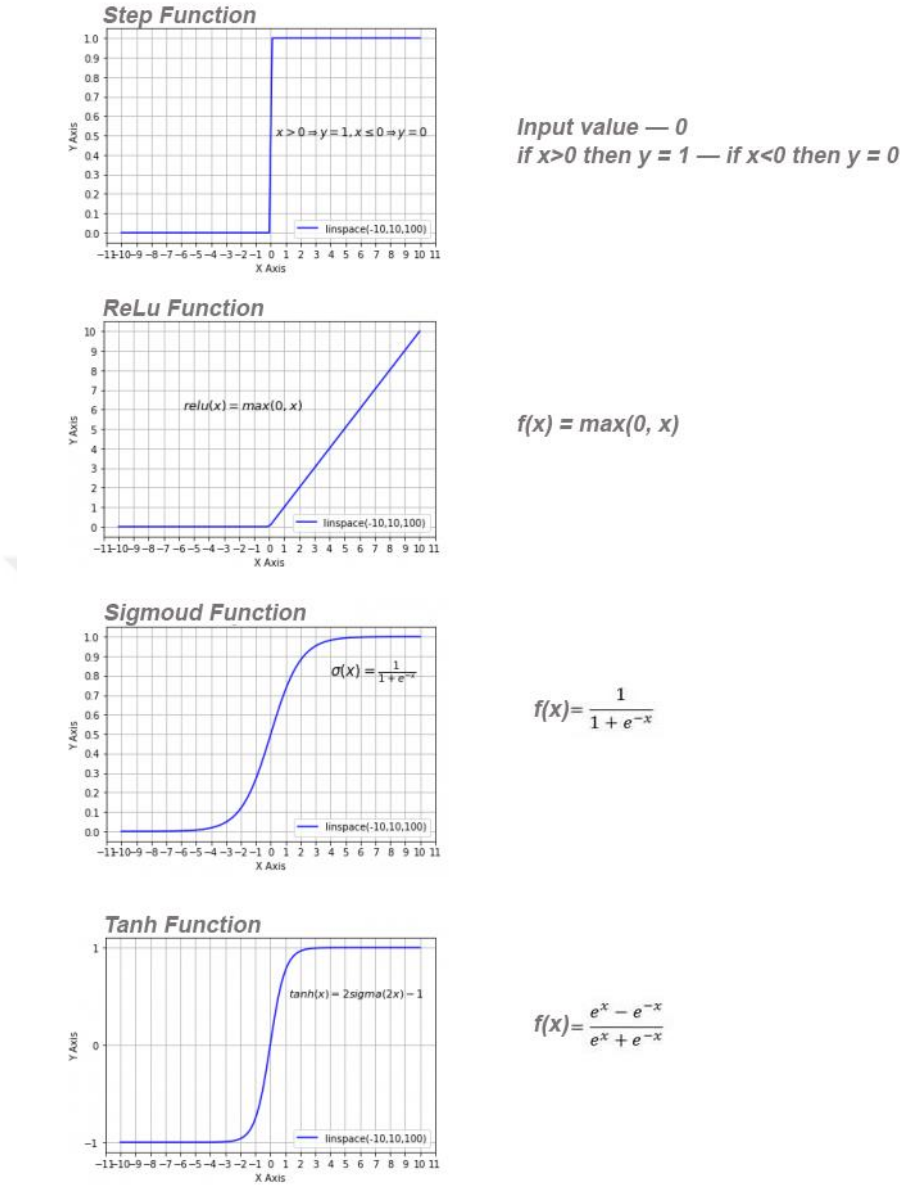


Figure 3.5. Commonly used ANN Activation Functions

5. Output (y): This is the value determined by the activation function. The output produced can be sent to another cell or to the outside of the world. In the case of feedback, the cell can use its own output value as input, with making feedback on itself. However, feedback can also be made to another cell. When shown in the form of a network, a cell appears to have more than one output. But this is for illustration purposes only. All of these outputs have the same value.

Artificial nerve cells come together to form an ANN. The ANN consists of three parts. These are the input layer, hidden layers, and the output layer. The cell numbers in the input and output layers are determined by the application. For example, an ANN to be installed for a 3-input 2-output system will have three cells in the input layer and two cells in the output layer. The number of hidden layers and the number of hidden cells in these layers are arbitrarily determined by the designer. As the number of hidden layers and hidden cell numbers increases, the fault tolerance of the ANN will increase, in addition to this the complexity of the operation and the duration of training will increase. The number of hidden layers and hidden cells is important for a good solution. The following figure shows an ANN for a 3-input 2-output system.

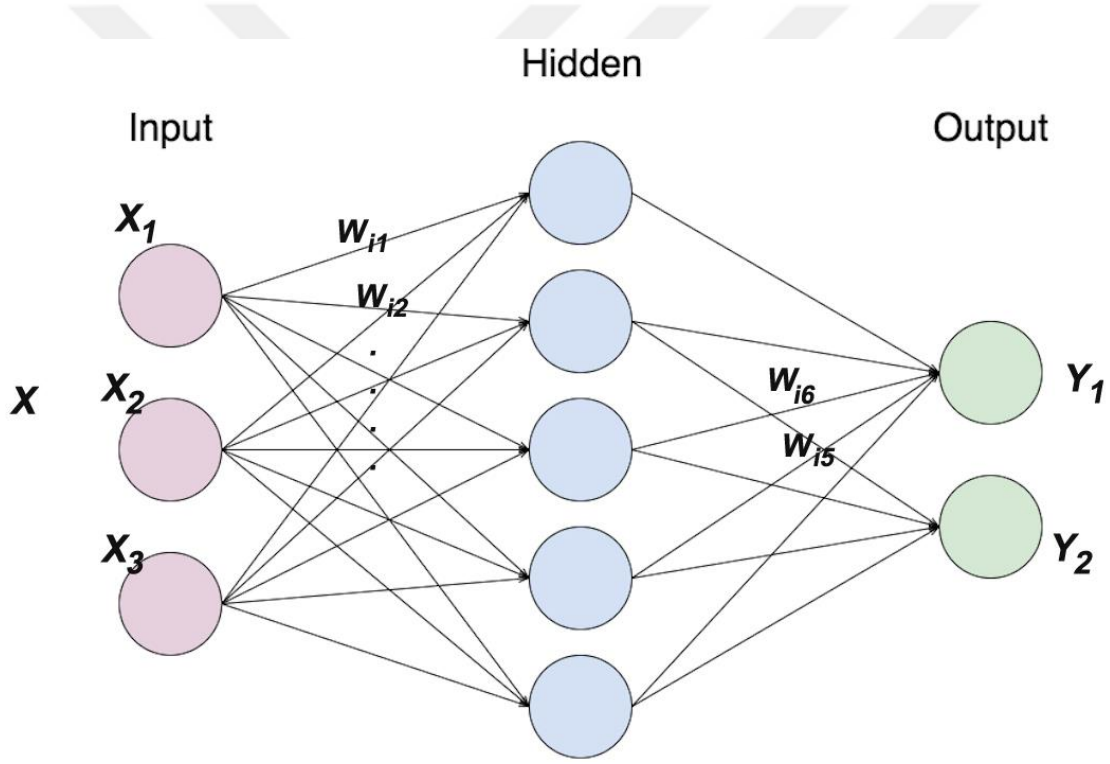


Figure 3.6. An example of a multilayer artificial neural network

For this system, one hidden layer and five hidden cells are used in this hidden layer. Weight values are indicated by w_{ij} notation. This notation indicates the link from i. cell to j. cell. Information in ANN is stored in the weight values of the connections. The input vector for the network shown above consists of x_1, x_2, and x_3. These inputs can be the voltage values given to the motor, the unit step input for a control system, numerical values representing the gray tones of an image or numerical values indicating a fingerprint. The

output vector consists of y_1 and y_2. In supervised learning algorithms, y_1 and y_2 are defined as target values. Initially, the weight values are arbitrarily determined, because even if the first iteration causes a major error, it must have a value. As examples are introduced into the artificial neural network, weights are updated and changed to produce the target value. The update process is performed according to the learning rules. When the error between the target values and the outputs produced by the ANN falls below a certain value, the training of the network is stopped and the test phase is taken to measure its performance. Generally, the test data is selected from samples that are not used during the training of the network. Weights do not change during the test phase. Using the weights from training, ANN produces output. The accuracy of these outputs gives information about the performance of the network. If the performance is considered adequate, it is assumed that the ANN is learned (Tsoukalas et al, 1996).

3.5.2 Learning in Artificial Neural Networks

In the ANN approach of the introduction of information is through examples. The examples represent the inputs and outputs of an event. It can be said that the ANN model, which learns the relationship between inputs and outputs, has gained the knowledge and experience to produce the outputs for the different inputs to be given. For example, if the response of a system to a unit step input is known, unit step input and output values and ANN models can be provided to learn the characteristics of the system. Now, the artificial neural network knows how the system will respond to other inputs. It is not necessary to use input and output values together always when creating samples. In some cases, examples can be generated with only input values. Depending on how the samples are formed, ANN models have three basic forms of learning.

1. Supervised Learning: Samples are created by using the input, and output set together. System inputs and the responses of the system to these inputs are introduced to the ANN, and the neural network is expected to learn the relations between the input, and output.

2. Reinforcement Learning: Samples are created with input values only. Instead of introducing the output values to the ANN, producing an output is expected by ANN, and the confirmative or rejection signal is sent to the ANN depending on whether the output is true or false.

3. Unsupervised Learning: There are no any teachers or supporters to help ANN models to learn. Only input values are introduced into the system. The ANN must be required to learn the relationships, similarities, and differences between these values by own. However, after the end of the learning process of the system, labeling that indicates what the outputs mean should be done by the user. This style of learning is mostly used for classification problems (Haykin, 1994).

According to this learning format, algorithms can be classified as Forward Feed models, Feedback models, and Race-based models.

3.5.2.1 Feedforward Neural Networks

Forward feed artificial neural networks allow for unidirectional signal flow. In addition, feedforward artificial neural networks are organized in many layers (Wilamowski, 2003). The following figure shows an example of a three-layer forward feed neural network. This network consists of an input layer, two hidden layers, and an output layer. Commonly used activation functions are shown in the section explaining the structure ANN models.

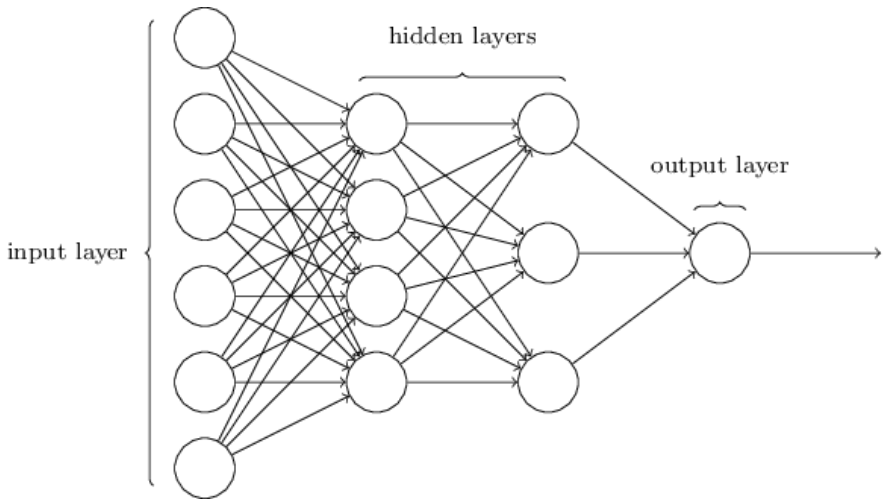


Figure 3.7. A three-layered forward feed neural network

In the forward feed neural network, the cells are arranged in layers and the outlets of the cells in a layer are introduced into the next layer via weights. The input layer transmits the information from external environments to the cells in the intermediate (hidden) layer without

changing it. The information is processed in intermediate and output layers for determining the network output. Thanks to this structure, forward feed networks perform a non-linear static function. It has been shown that the forward feed 3-layered ANN can approach any continuous function to the desired accuracy if there are sufficient numbers of cells in the middle layer. The backward propagation algorithm, which is the most known algorithm, is used effectively in the training of such neural networks. Both samples and outputs, which should be obtained from samples (expected outputs) are given to the network. Inferring from these examples, the network creates a solution area showing the problem area. Forward feed models include multilayer perceptron.

3.5.2.2 Multilayer Perceptron (MLP)

As a result of perceptron and Adaline methods are not able to produce non-linear solutions, multilayer perceptron networks that are better in both architecture and education algorithm were offered. The most widely used model of ANN is the multilayer perceptron network (MLP). These networks are capable of providing solutions to 95% of engineering problems particularly. MLP networks emerged as a result of efforts to solve the XOR problem. These networks consist of 3 layers; Input layer, Intermediate layers, Output layer (Bulsari, 1995).

The numbers of process elements in the input and output layers are decided by looking at the problem. There is no any method that indicates the number of intermediate layers and the number of process elements in each intermediate layer. This is determined by trial and error (Rumelhart et al., 1986). All elements in the input layer are linked to elements in the hidden layer. They also connected to all process elements in the output layer. The information flow is toward from the input layer to the intermediate layer, then from intermediate layer to output layer. The training of the MLP network occurs according to the generalized delta rule. Because of MLP networks use supervised learning strategies, both inputs and outputs that the network must produce in response to those inputs are shown to the network during training. The used philosophy of the learning rule is the minimization of the difference (error) that is between the outputs produced by the network and the expected outputs with distribution to the weights of the network over time during training. The inputs are first presented to the network and the output from these inputs is generated during the learning. This process is called forward calculation. Then the output produced is compared with the expected output and the

error is distributed backward and the weights are changed. This is called backward calculation.

3.5.3 The Hyperparameters of ANN Models

The selection of appropriate hyperparameters ensures that ANN training to be performed at a lower cost and with higher performance. The hyperparameters of ANN models optimized in this thesis study are listed below.

1. Number of epochs

While the model is being trained, not all of the data does not participate in training at the same time. They are included in the training in a certain number and in pieces. The first part is trained, the performance of the model is tested, it is updated weights by backpropagation according to the success. Then the model is re-trained with the new training set and the weights are updated again. This process is repeated in each training step and the most suitable weight values are calculated for the model. Each of these training steps is called "epoch". Since the optimal weight values to solve the problem are calculated step by step, the performance in the first epochs will be low and the performance will increase as the epoch number increases. However, after a certain step, the learning status of the model will be considerably decreased. The size of the epoch number also varies according to the type of problem. The high number of epochs does not mean that the network will produce higher accuracy results. You et al. (2017), regardless of the epoch number and other hyperparameter values, have observed that the accuracy gain is slower after one point and accuracy does not increase more.

2. Mini-Batch Size

In deep learning applications, the learning process of all the data in the data set at the same time is not efficient in terms of time and memory. Because, in each iteration of learning, gradient descent calculation is performed with the back propagation process and the weight values are updated in this way. In this calculation, the numbers of data and the calculation processes have a direct proportion. To solve this problem; the data set is divided into small groups and the learning process is performed on these selected small groups. In this way, the processing of multiple inputs into pieces is called "mini-batch". When designing the model,

the value specified as the mini-batch parameter means that how many data will process by the model at the same time.

When selecting the mini-batch value, the optimal value should be determined between 1 and the number of all data in the training set, ie it should be neither too small nor too large. This will provide a quick learning process. Another criterion of batch size is memory size. Memory size and batch size show a parallel situation, in a low memory environment, keeping batch size large is problematic. Therefore, it will be efficient to calculate the maximum batch value to be used before designing the model. In addition, the small size of the batch creates a regularization effect, and when the data is given to the model in large batches, overfitting may be more. When performing the batch operation, the data set is divided into parts according to the value determined as batch value and the training of the model is performed on this part at each iteration. In some cases, however, it may be the data grouped within itself. This will create a correlation within the data set; the test set to be selected from this data set will also provide high performance, thus it will be overfitting. To prevent this, the data set must be shuffled before the data set is divided into pieces. In batch selection, it is important to randomly select data.

3. Activation Functions

Activation functions are hyperparameters which play an important role in the success of the model in ANN models. Activation functions add non-linearity to the model. In the hidden layer, in the linear function y = f (x, w), the matrix is multiplied and the output is converted to a non-linear value after the weight of the neurons is calculated. Because deep learning methods are more effective than other methods in solving non-linear problems, the problem that is tried to be solved by deep learning methods is generally a non-linear problem. The conversion of the value obtained as a result of matrix multiplication to non-linear is done by activation functions. Activation functions are used for non-linear transformation processes in multi-layer artificial neural networks. In order to calculate the gradient descent from hidden layers, the output of the hidden layers is normalized by some activation functions. Commonly used activation functions were mentioned in the previous section. These functions are included in the list of parameters to be selected for the model and can be optimized. Choosing the appropriate function makes a significant contribution to the success of the model.

4. Dropout

Srivastava et al. (2014) interpreted dropout as a way to regulate an ANN by adding noise to its hidden units and stated that it is a technique that reduces overfitting to improve ANN models. Dropout is the technique of removing certain nodes from the hidden or input layer according to certain rules.

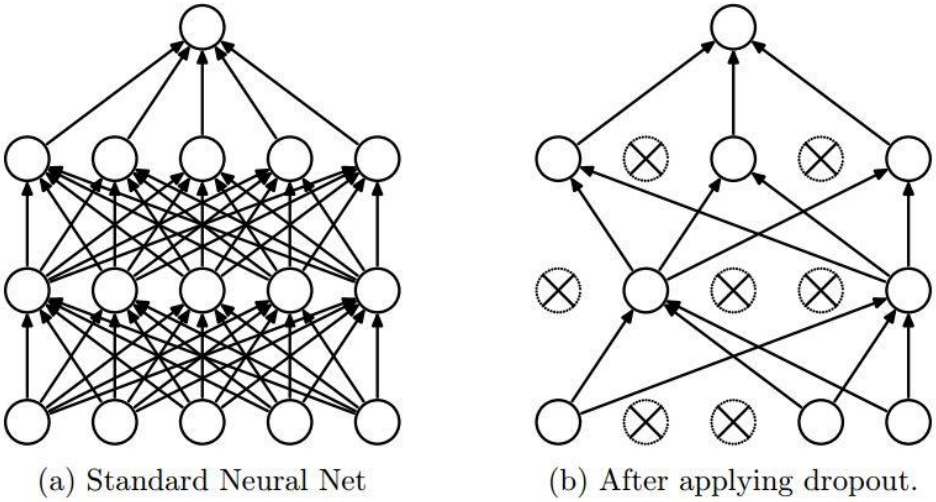


Figure 3.8. An example of ANN with and without dropout

Dropout technique is usually used after fully-connected layers. Using dropout, the links on fully-connected layers are broken. Thus, the nodes have less information about each other, and as a result, the nodes are less affected by each other's weight changes. Therefore, more robust models can be created with dropout method. At the same time, better learning will take place due to different hidden unit combinations in each layer work with each other and performance will increase. Dropout value is usually used as 0.5. It is also widely used in different values. It varies according to the problem and data set. When the dropout value is used as a threshold value, it is defined as a value in the range [0, 1].

3.6 k-Fold Cross Validation

Cross-validation or "k-fold cross validation" is the process of dividing the data set into random "k" groups. One of the groups is used as a test set and the rest is used as a training set. The hold-out method is the process of dividing the data set into "training" and "test" sets. When using the hold-out method, it commonly uses 80% of the data for training and 20% of

the data for testing, and this ratio varies with the size of the data set. The cross validation method is one of the most preferred methods. Since both test and train operations are performed on all data, it will be positively effective for the success of the model. The hold-out method is useful in a very large data set and it takes less time. Because cross validation uses multiple training-test partitions, it takes more time than hold-out as it requires more computational power.

Stone (1974) and Geisser (1975) used cross validation to select appropriate model parameters instead of using cross-modeling to select appropriate model parameters. Weiss et al. (1991) proposed 10 fold cross validation (k=10) procedures to check the generalization ability of the model in data mining where the data set is small.

In cross validation, the purpose of separating the data set as a training and test set is to avoid possible overfitting and to understand how the model performs on a data set that it has not seen before. However, there may be some errors due to distribution during the training and testing phase of the model. In order to minimize these errors, k-fold cross validation technique is used. The training data set is randomly divided into k parts. k-1 piece is used for training, one piece is used for test set and this process is repeated k times. The values obtained in each round are summed up; it is averaged and the performance of the model is evaluated.

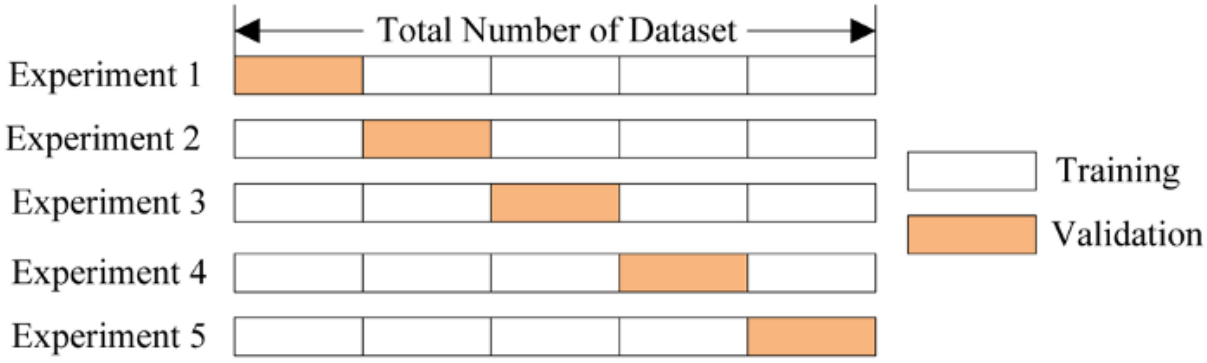


Figure 3.9. An example of 5-fold cross validation

In the cross validation process performed as above, the data set is divided into five groups and the model is trained and tested five times, so that each group will have the chance to become a test set.

4. RESULTS AND DISCUSSION

The data set of this study consists of the properties and price information of the houses, which are announced on the Internet between February and July 2018 for the districts of Istanbul. The data set size and descriptive statistics of each district are shown in the table below.

Table 4.1. Descriptive statistics of districts

District	Esenyurt	**Size of Data Set**	24.224			
Mean	***StDev***	***Variance***	***Coef Var***	***Sum***	***Min***	***Max***
244.602	141.351	19.980.205.260	58	5.925.234.816	17.000	3.850.000
District	Kadıköy	**Size of Data Set**	21.579			
Mean	***StDev***	***Variance***	***CoefVar***	***Sum***	***Min***	***Max***
1.347.292	1.588.141	2.522.190.000.000	118	29.073.222.809	40.000	33.100.000
District	Maltepe	**Size of Data Set**	20.321			
Mean	***StDev***	***Variance***	***CoefVar***	***Sum***	***Min***	***Max***
461.048	283.671	80.469.290.285	62	9.368.960.152	25.000	8.539.200
District	Beylikdüzü	**Size of Data Set**	19.729			
Mean	***StDev***	***Variance***	***CoefVar***	***Sum***	***Min***	***Max***
343.146	218.147	47.588.160.294	64	6.769.918.581	25.000	4.447.500
District	Ümraniye	**Size of Data Set**	17.407			
Mean	***StDev***	***Variance***	***CoefVar***	***Sum***	***Min***	***Max***

422.752	265.732	70.613.585.826	63	7.358.846.468	25.000	7.709.000

District	Bahçelievler	**Size of Data Set**	16.488

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
403.808	329.964	108.876.000.000	82	6.657.983.984	30.000	7.649.700

District	Pendik	**Size of Data Set**	16.161

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
348.784	177.629	31.552.034.927	51	5.636.697.116	33.500	3.940.000

Table 4.1 (Continued)

District	Eyüp	**Size of Data Set**	13.414

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
601.788	844.113	712.527.000.000	140	8.072.387.500	30.000	17.790.000

District	Kartal	**Size of Data Set**	11.657

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
437.242	371.252	137.828.000.000	85	5.096.934.951	99.000	13.650.000

District	Avcılar	**Size of Data Set**	10.367

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
320.381	224.979	50.615.680.811	70	3.321.385.161	25.000	4.104.400

District	Ataşehir	**Size of Data Set**	9.440

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
666.896	562.611	317.000.000.000	84	6.295.500.263	50.000	7.116.000

District	Çekmeköy	**Size of Data Set**	9.221

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
352.907	250.716	62.858.717.707	71	3.254.154.752	25.000	5.930.000

District	Üsküdar	**Size of Data Set**	8.395

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
966.627	1.463.469	2.141.740.000.000	151	8.114.833.410	90.000	17.874.000

District	Bağcılar	**Size of Data Set**	8.141

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
384.333	228.451	52.189.874.802	59	3.128.854.161	59.000	5.250.000

District	Tuzla	**Size of Data Set**	6.985

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
354.947	207.854	43.203.234.745	59	2.479.301.852	50.000	5.633.500

District	Kağıthane	**Size of Data Set**	6.807

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
388.464	189.063	35.744.767.795	49	2.644.276.244	49.000	3.900.000

District	Gaziosmanpaşa	**Size of Data Set**	6.515

Table 4.1 (Continued)

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
371.948	200.246	40.098.370.78	54	2.423.239.520	25.000	2.500.000

District	Şişli	**Size of Data Set**	5.829

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
606.412	462.049	213.000.000.000	77	3.534.777.287	50.000	6.750.000

District	Bakırköy	**Size of Data Set**	5.697

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
2.229.823	2.846.243	8.101.100.000.000	128	12.703.303.179	50.000	41.510.000

District	Beşiktaş	**Size of Data Set**	5.421

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
4.048.751	5.821.321	33.887.800.000.000	144	21.948.278.034	50.000	83.020.000

District	Büyükçekmece	**Size of Data Set**	5.153

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
474.641	401.307	161.047.000.000	85	2.445.827.602	25.000	6.819.500

District	Sarıyer	**Size of Data Set**	4.589

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
2.772.216	3.650.166	13.323.700.000.000	132	12.721.699.471	80.000	41.510.000

District	Fatih	**Size of Data Set**	4.217

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
444.117	296.469	87.893.763.688	67	1.872.839.502	25.000	3.707.200

District	Esenler	**Size of Data Set**	4.147

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*

262.987,92	127.986,31	16.380.494.521,31	48,67	1.090.610.884,00	45.000,00	1.599.000,00

District	Güngören	**Size of Data Set**	3.940

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
364.531	219.295	48.090.373.459	60	1.436.251.151	60.000	3.498.700

District	Arnavutköy	**Size of Data Set**	3.598

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
250.488	86.792	7.532.826.626	35	901.256.254	76.000	1.110.000

Table 4.1 (Continued)

District	Bayrampaşa	**Size of Data Set**	3.309

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
445.004	234.914	55.184.770.557	53	1.472.517.836	95.000	3.500.000

District	Zeytinburnu	**Size of Data Set**	3.200

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
602.237	611.407	373.818.000.000	102	1.927.158.895	37.000	11.860.000

District	Silivri	**Size of Data Set**	2.715

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
292.893,27	173.701,29	30.172.137.376,44	59,31	795.205.232,00	25.000,00	4.151.000,00

District	Sultanbeyli	**Size of Data Set**	2.561

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
303.419	150.814	22.744.741.852	50	777.057.055	40.000	2.350.000

District	Beyoğlu	**Size of Data Set**	2.440

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
900.749	1.780.740	3.171.040.000.000	198	2.197.828.678	60.000	22.534.000

District	Beykoz	**Size of Data Set**	818

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
1.197.941	1.826.753	3.337.030.000.000	152	979.915.800	45.000	23.720.000

District	Şile	**Size of Data Set**	517

Mean	*StDev*	*Variance*	*CoefVar*	*Sum*	*Min*	*Max*
349.778	102.493	10.504.908.759	29	180.835.000	90.000	675.000

District	Çatalca	**Size of Data Set**	480

Mean	StDev	Variance	CoefVar	Sum	Min	Max
318.405	97.671	9.539.688.155	31	152.834.390	135.000	800.000

District	Adalar	**Size of Data Set**	274

Mean	StDev	Variance	CoefVar	Sum	Min	Max
1.147.843	2.116.296	4.478.710.000.000	184	314.509.060	185.000	33.100.000

The total size of the data set used in the study is 302.050 and the largest data set is in the district of Esenyurt with approximately 24.000. It is followed by Kadıköy, Maltepe districts with approximately 21.000 and 20.000. The districts with the lowest data are; Adalar (274), Çatalca (480), Şile (517). Beşiktaş (83.020.000 TL), Sarıyer (41.510.000 TL) and Bakırköy (41.510.000 TL) districts are among the districts with the highest sales prices. The districts have the lowest house price; Esenyurt (17.000 TL), Maltepe, Beylikdüzü, Ümraniye, Fatih, Gaziosmanpaşa, Çekmece, Büyükçekmece, Avcilar, Silivri (25.000 TL) are the districts. Beşiktaş, which has the highest variance with "33.887.800.000.000", has been observed with a standard deviation value of "5.821.321". The Sarıyer district which has a variance of "13.323.700.000.000" and a "3.650.166" standard deviation value came second.

All data and qualifications related to the dwellings belonging to these districts were used as a database for the thesis study. The SQL Server program was used for data (unprocessed). Microsoft SQL Server is the most used database server software. It is a relational database management system that enables the creation and management of databases. With the relational database system, the tables in the database are linked to each other according to the relationship between them. In the first step in this study, the data is stored in the related tables on the SQL system. The data set contains the properties that the houses can have and the values of each house in this property.

4.1 Data Preprocessing Steps

In the second step of the study, the data were preprocessed. The data preprocess aims to make the data in the database or data warehouses where there is a lot of data, statistically healthy before the analysis phase. In the second step of the study, the data were preprocessed. The data preprocess aims to make the data that is in the database or data warehouses where there is a lot of data, is statistically healthy before the analysis phase. To make the data

healthy necessitates finding the data, which contains deficient, inadequate, inconsistent, and contrarian features, then finding the solutions to those problems with appropriate methods. During data entry or transmission, missing, excess, repetitive data may be entered. Therefore, the data is very important in our preprocessing study. For this reason, a serious study was performed on the data before the analysis. First, the most likely values for the inconsistent data that are missing on the data transferred from the database (SQL) have been used or cleared. Data merge is performed for the property values of the data from different tables (SQL Server). After checking the combined data, appropriate corrections were made for inconsistent records.

Data conversion was performed on the data in the third step of the study. The fact that the original forms of the data are different from each other because of various reasons is inevitable in databases. With data conversion, the data are brought into appropriate forms to be statistically meaningful and integrated before the analysis. The total number of qualifications of the data in our study is 157 and there is no reduction in these qualifications before the analysis. These qualifications are listed in the following tables.

Table 4.2. Main features of houses in the data set

MAIN FEATURES
1. The Price
2. The Neighborhood
3. The Apartment's area (meter)
4. The Number of rooms
5. The Building's age
6. The Houses' floor
7. The Number of floors
8. The Heating type
9. The Number of bathrooms
10. Usage status

Table 4.3. Dummy Variable in the data set

DUMMY VARIABLE (IT GETS 0 OR 1 VALUE)	
Facade	
11. East	12. West
13. North	14. South
Interior Features	
15. ADSL	16. Woodwork
17. Smart Home	18. Alarm (Burglar)
Table 4.3. (Continued)	
19. Alarm (Fire)	20. Alaturka Toilet
21. American Kitchen	22. Built-in Oven
23. Elevator	24. Balcony
25. Barbecue	26. White Goods
27. Painted	28. Dishwasher
29. Refrigerator	30. Wall Paper
31. Shower	32. Parent Bathroom
33. Fiber Internet	34. Owen
35. Dressing Room	36. Inbuilt Cupboard
37. Entryphone	38. Hilton Bathroom
39. Intercom System	40. Thermopane
41. Jacuzzi	42. Papier-Mache
43. Storeroom	44. Air Conditioning
45. Bathtub	46. Laminate Floor
47. Marley	48. Furniture
49. Kitchen (Built-in)	50. Kitchen (Laminate)
51. Kitchen Natural Gas	52. PVC Joinery
53. Blinds	54. Hardwood Flooring
55. Sauna	56. Ceramic Floor
57. Countertop Stove	58. Spot Lighting
59. Terrace	60. Cloakroom
61. Wi-Fi	62. Washing Machine
63. Laundry Room	
Exterior Features	
64. Steel Door	65. Water Heater

66. Fireplace	67. Elevator
68. Security	69. Hydrophore
70. Heat Insulation	71. Generator
72. Cable TV	73. Indoor Parking
74. Doorman	75. Kindergarten
76. Car Park	77. Playground
78. Sound Insulation	79. Siding
80. Playfield	81. Reservoir
82. Tennis Court	83. Satellite
Table 4.3. (Continued)	
84. Fire Escape	85. Swimming Pool
86. Indoor Pool	
Suitable for Disabled	
87. Car Parking	88. Elevator
89. Bathroom	90. Large Corridor
91. Entry / Ramp	92. Stairs
93. Kitchen	94. Room Door
95. Park	96. Socket / Electrical Switch
97. Handrail / Railing	98. Toilet
99. Swimming Pool	
Environment	
100. The Mall	101. Municipality
102. Mosque	103. Cemevi
104. Beachfront	105. Pharmacy
106. Amusement Center	107. Fair
108. Hospital	109. Synagogue
110. Church	111. High School
112. Market	113. Park
114. Police Station	115. Health Clinic
116. District Bazaar	117. Gym
118. University	119. Primary Education
120. Fire Department	121. Town Center
Transportation	

122.	Highway	123.	Eurasian Tunnel
124.	Bosphorus Bridges	125.	Street
126.	Sea Bus	127.	Dolmush
128.	E-5	129.	Airport
130.	Marmaray	131.	Metro
132.	Metrobus	133.	Minibus
134.	Bus Stop	135.	Beach
136.	TEM	137.	Telpher
138.	Tramway	139.	Railway Station
Table 4.3. (Continued)			
140.	Wharf		
View			
141.	Bosphorus	142.	Sea
143.	Nature	144.	Lake
145.	Pool	146.	City
Housing Type			
147.	Mezzanine	148.	Mezzanine Dublex
149.	Garden Duplex	150.	Garden Floor
151.	Having A Garden	152.	Top Floor
153.	Floor Duplex	154.	Detached House (Entry)
155.	Reverse Duplex	156.	Triplex
157.	Roof Duplex		

The qualifications of the data that must be converted are listed below.

1) The Neighborhood: The neighborhood areas of each district are digitized such as starting from 1 and increasingly.

2) The Houses's floor: This property contains numerical and text values.
   i. Villa Type
   ii. Entrance floor
   iii. Garden Floor
   iv. High entrance

v. Self-contained
vi. Penthouse
vii. Basement
viii. Numerical Values (1,2,3…)

3) The Heating type: Since this property does not contain numerical values, it has been digitized as increasing starting from 1.
   i. No Heating
   ii. Air conditioning
   iii. Fireplace
   iv. Stove
   v. Natural Gas Stove
   vi. Natural Gas (Combi boiler)
   vii. Room heater
   viii. Floor Heating
   ix. Central heating

4) The Building's age: This qualification contains numerical and text values. For example, the data that has a value of 5 to 10 is available.

5) Usage status: It includes "Empty", "Tenant", and "Owner"

In the study, these properties were digitized using the Python program. The data of each district was used as a separate data set, and the digitization process was applied to each data set. The puppet variables in the set of variables are the variables that receive the value 1 if the dwelling has the specified property, otherwise, it receives the value 0. For example, the "elevator" variable is a puppet variable and is expressed as 1 if there is, and 0 if there is no. The variables in the study are puppet variables except the price, the neighborhood, the houses' floor, the number of rooms, the heating type , the number of baths, the building's age, and the apartment's area (square meters).

In the second stage of the data conversion, the digitized input data in the [0, 1] range was normalized using the python program. The Min-Max Normalization function is used. In

this method, the largest and smallest values in the data set are determined for calculation. In the second stage, normalization is performed on all data based on these values. The goal is to normalize the smallest value to 0 and the maximum value to 1, and then to spread all the data in this range of 0 -1. The formula for this function is listed below.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

Here, X is the value that the qualification takes at that moment. X_{min} and X_{max} are the smallest and largest values that the qualification takes.

The price value as an output value is normalized in [0,1] and [-1,1] ranges then compared. An example of a normalized district (Pendik) data set is shown in the following figure. It is determined that the estimations made by normalizing in the range of [-1,1] are closer to reality. It was observed that the estimations were better when normalized in the range [0, 1] in Çatalca and Şile districts. As a result, a normalization has been elaborated for these districts.

Table 4.4. An example of a normalized (Pendik) district's data set

	Price [-1,1]	Price [0,1]	Real Price
1	-0,962370408	0,018814796	107000,00
2	-0,965442212	0,017278894	101000,00
3	-0,945987457	0,027006272	139000,00
4	-0,935236145	0,032381928	160000,00
5	-0,96390631	0,018046845	104000,00
6	-0,963394343	0,018302829	105000,00
7	-0,963394343	0,018302829	105000,00
8	-0,962370408	0,018814796	107000,00
9	-0,963394343	0,018302829	105000,00
10	-0,937795981	0,031102009	155000,00
11	-0,930116473	0,034941764	170000,00
12	-0,930116473	0,034941764	170000,00
13	-0,962370408	0,018814796	107000,00
14	-0,961858441	0,019070779	108000,00

	....	….	…..
	….	….	…..
16162	-0,786765647	0,106617177	450000,00

4.2 Feature Selection with SelectKBest

After the data conversion step, the specified number of features were selected using the filtering method, which is one of the feature selection methods. The best features were determined with "SelectKBest", one of the filtering methods of the SciKit-Learn library, a Python program library. With this method, several properties that can vary according to districts were selected to be used in the input layer of the ANN model. In the SelectKBest method, scores were calculated using two different score functions (chi2, f_regression), and the features to be used for the model were selected according to these scores. The feature selection processes were applied six times and in different numbers for each district. Three different numbers of features (such as 20, 50, 100) to be selected using chi2 and f_regression score functions were determined. When the results are examined, the selected features of districts differ according to data sets and score functions. In the tables below, the selected features of Esenyurt (district with the largest data) and the selected features of Kartal (district having the best prediction performance) are shown as examples.

Table 4.5. The selected features of Esenyurt by SelectKBest method

District: Esenyurt	**Score Function : chi2**	**Normalization [-1,1]**	**20 Features**
1. The Apartment's area 2. The Heating type 3. The Houses's floor 4. Usage status 5. The Number of floors 6. The Neighborhood 7. Parent Bathroom 8. Garden Duplex 9. Kindergarten 10. Pool		11. The Building's age 12. Owen 13. Dishwasher 14. Air Conditioning 15. indoor Parking 16. Garden Floor 17. Dressing Room 18. Water Heater 19. Having a garden 20. inbuilt Cupboard	
District:	**Score Function :**	**Normalization [-1,1]**	**20 Features**

Esenyurt	f-regression		
1. The Number of rooms 2. Dishwasher 3. The Building's age 4. Built-in Oven 5. The Number of bathrooms 6. The Number of floors 7. Satellite 8. The Heating type 9. Parent Bathroom 10. Owen		11. Fire Escape 12. The Apartment's area 13. Hydrophore 14. Air Conditioning 15. Kindergarten 16. Generator 17. Siding 18. indoor Parking 19. Garden Floor 20. Sound insulation	

In the table above, selected features as a result of two different feature selection process applied to Esenyurt district are shown. As a result of the SelectKBest method applied, the first 20 features selected differed according to the score functions. The feature with the best score according to chi2 function is "The Apartment's area", whereas the feature with the best score according to f_regression function is "The Number of rooms". It is observed that "The Number of rooms" feature is not among the top 20 features as a result of feature selection made according to chi2 function, and "The Number of floors" feature is in the first place in both feature selection. The heating type feature is ranked second in the selection of features using the chi2 function, and it is ranked 8th position according to the f_regression function.

Table 4.6. The selected features of Kartal by SelectKBest method

District: Kartal	Score Function : chi2	Normalization [-1,1]	50 Features
1. The Apartment's area 2. The Houses' floor 3. The Heating type 4. Woodwork 5. The Number of floors 6. The Neighborhood 7. Bosphorus	18. Floor Duplex 19. Water Heater 20. Pool 21. Playfield 22. Mezzanine Dublex 23. Jacuzzi 24. Swimming Pool 25. Wall Paper 26. Dishwasher	35. Smart Home 36. American Kitchen 37. Beachfront 38. Barbecue 39. Refrigerator 40. Swimming Pool(Disabled) 41. The Building's age 42. Triplex	

8. Marley 9. Terrace 10. Having a garden 11. Parent Bathroom 12. Washing Machine 13. Furniture 14. Tennis Court 15. Garden Duplex 16. Spot Lighting 17. Owen	27. Telpher 28. inbuilt Cupboard 29. Storeroom 30. Alarm(Burglar) 31. Dressing Room 32. Playground 33. Detached House(Entry) 34. Reverse Duplex	43. Lake 44. Garden Floor 45. Bathtub 46. Tramway 47. Air Conditioning 48. Blinds 49. Laundry Room 50. White Goods

District: Kartal	Score Function : f-regression	Normalization [-1,1]	50 Features

Table 4.6. (Continued)

1. The Number of rooms 2. ADSL 3. The Number of bathrooms 4. The Building's age 5. Usage status 6. Kindergarten 7. Intercom System 8. Ceramic Floor 9. Sea Bus 10. The Number of floors 11. Garden Floor 12. Fiber Internet 13. Hydrophore 14. Swimming Pool 15. Smart Home 16. Air Conditioning 17. Playground	18. Dishwasher 19. Satellite 20. Diafon 21. Cloakroom 22. Wharf 23. Spot Lighting 24. Reservoir 25. Sea 26. The Heating type 27. Sound insulation 28. Municipality 29. Beachfront 30. indoor Pool 31. Gym 32. Beach 33. Parent Bathroom 34. Alarm(Fire)	35. Generator 36. Siding 37. Reverse Duplex 38. Doorman 39. Detached House(Entry) 40. Alarm(Burglar) 41. indoor Parking 42. Countertop Stove 43. Fire Escape 44. Flat area (meter) 45. Kitchen(Built-in) 46. Elevator(Exterior) 47. Built-in Oven 48. Alaturka Toilet 49. Wi-Fi 50. Railway Station

When Esenyurt and Kartal districts are compared, it is remarkable that the features in the first order are similar according to the result of feature selection using chi2 function. In this district, the "The Number of rooms" feature is the first feature obtained with the f_regression function. The “The Neighborhood” feature is among the features selected in the

results obtained with the chi2 function as in Esenyurt district. However, this feature is not among the features selected by f_regression of both districts.

4.3 Determination of optimal ANN Parameters by using Grid Search

In this thesis, Grid Search method was used to determine the best model parameters before the ANN model was applied to district data sets. The model parameters were optimized for each district using the GridSearchCV method of the scikit-learn library of Python. Furthermore, 5-fold cross-validation was combined with this method. When the results of the Grid Search applied to each district data set were examined, it was observed that some parameters were the same and others were different. The parameter with the smallest mean squared error (MSE) value was considered for parameter selection. It has been observed that it takes quite a long time to determine the parameters for districts such as Esenyurt, Kadıköy, and Maltepe, which have large data sets. The parameter ranges to be used for GridSearh have been determined intuitively and with consideration of previous analysis.

The Parameters' ranges are below:

Epoch = [100, 200, 500]

Batch-size = [60, 100]

Activation Function = ["relu", "sigmoid", "tanh"]

Dropout = [0.2, 0.3, 0.4, 0.5]

Apart from these parameters, one of the other model parameters, optimization algorithm parameter was observed to be generally "Adam" and it was selected as the model parameter. According to the districts, ANN has been trained for the combination of all values in these ranges and the most suitable parameters have been tried to be determined as a result of observation.

Table 4.7. The parameters values of Kadıköy district after applying Grid Search method

Kadıköy (Size of data set : 21.579)			
Time : 4 hour	**Mean**	**Std**	**Test Results**

Epochs and Batch Size	0.003447 (0.000443) with: {'batch_size': 60, 'epochs': 100} **0.003197 (0.000342) with: {'batch_size': 60, 'epochs': 200}** 0.003474 (0.000489) with: {'batch_size': 60, 'epochs': 500} 0.003367 (0.000359) with: {'batch_size': 100, 'epochs': 100} 0.003199 (0.000360) with: {'batch_size': 100, 'epochs': 200} 0.003205 (0.000308) with: {'batch_size': 100, 'epochs': 500}
200-60	
Activation function (input and hidden layer)	0.003511 (0.000568) with: {'activation': 'relu'} 0.003405 (0.000400) with: {'activation': 'tanh'} **0.003323 (0.000357) with: {'activation': 'sigmoid'}**
sigmoid	
Activation function (output layer)	0.857313 (0.002086) with: {'activation': 'relu'} **0.003406 (0.000214) with: {'activation': 'tanh'}** 0.857313 (0.002086) with: {'activation': 'sigmoid'}
tanh	

Table 4.7. (Continued)

Dropout	**0.003233 (0.000380) with: {'dropout_rate': 0.2}** 0.003316 (0.000408) with: {'dropout_rate': 0.3} 0.003630 (0.000403) with: {'dropout_rate': 0.4} 0.003991 (0.000554) with: {'dropout_rate': 0.5}
0.2	

Table 4.8. The parameters values of Avcılar district after applying Grid Search method

Avcılar (Size of data set : 10.367)			
Time : 3 hour	**Mean**	**Std**	**Test Results**
Epochs and Batch Size	0.005619 (0.001301) with: {'batch_size': 60, 'epochs': 100} 0.005164 (0.000917) with: {'batch_size': 60, 'epochs': 200} **0.004746 (0.000901) with: {'batch_size': 60, 'epochs': 500}** 0.006027 (0.001296) with: {'batch_size': 100, 'epochs': 100} 0.005199 (0.001353) with: {'batch_size': 100, 'epochs': 200} 0.005113 (0.001330) with: {'batch_size': 100, 'epochs': 500}		
500-60			
Activation function (input and hidden layer)	**0.005611 (0.001289) with: {'activation': 'relu'}** 0.006135 (0.001187) with: {'activation': 'tanh'} 0.006956 (0.001515) with: {'activation': 'sigmoid'}		
relu			
Activation function (output layer)	0.743505 (0.001389) with: {'activation': 'relu'} **0.005525 (0.001243) with: {'activation': 'tanh'}** 0.743505 (0.001389) with: {'activation': 'sigmoid'}		
tanh			
Dropout	**0.005397 (0.001291) with: {'dropout_rate': 0.2}**		

0.2	0.005536 (0.001257) with: {'dropout_rate': 0.3} 0.005878 (0.001447) with: {'dropout_rate': 0.4} 0.006327 (0.001315) with: {'dropout_rate': 0.5}

Table 4.9. The parameters values of Şişli district after applying Grid Search method

Şişli (Size of data set : 5.829)			
Time : 2 hour	**Mean**	**Std**	**Test Results**
Epochs and Batch Size	0.018861 (0.001038) with: {'batch_size': 60, 'epochs': 100} 0.018996 (0.001113) with: {'batch_size': 60, 'epochs': 200} 0.019534 (0.001072) with: {'batch_size': 60, 'epochs': 500}		
Table 4.9. (Continued)			
200-100	0.019003 (0.001070) with: {'batch_size': 100, 'epochs': 100} **0.018830 (0.001030) with: {'batch_size': 100, 'epochs': 200}** 0.018892 (0.001075) with: {'batch_size': 100, 'epochs': 500}		
Activation function (input and hidden layer)	0.018853 (0.001039) with: {'activation': 'relu'} 0.018790 (0.001059) with: {'activation': 'tanh'} **0.018772 (0.001095) with: {'activation': 'sigmoid'}**		
sigmoid			
Activation function (output layer)	0.714421 (0.003563) with: {'activation': 'relu'} **0.018786 (0.001093) with: {'activation': 'tanh'}** 0.714443 (0.003567) with: {'activation': 'sigmoid'}		
tanh			
Dropout	0.018801 (0.001110) with: {'dropout_rate': 0.2}		
0.3	**0.018787 (0.001093) with: {'dropout_rate': 0.3}** 0.018801 (0.001095) with: {'dropout_rate': 0.4} 0.018870 (0.001095) with: {'dropout_rate': 0.5}		

In the tables above, the results obtained from the districts of Kadıköy, Avcılar and Şişli are shown after applying the Grid Search method. The optimization process for each model parameter is discussed in detail in the next section.

4.3.1 The Parameters Optimized with Grid Search Method

Using the Grid Search method, the model parameters were optimized in an appropriate order to the ANN model. In the first step, the best parameters were selected by determining the appropriate parameter range for Epoch and Batch-Size. Secondly, the most common activation functions for input and hidden layer were added to the parameter list and the most suitable parameter was found. In the third step, the best activation function was determined for the output layer. Finally, for Dropout value, the best value was selected among the values in the specified range. For all steps, the appropriate parameter setting process could be performed with the Grid Search method at one time. However, it is highly probable that there will be too many combinations for the five parameter groups and it will take a long time and there may be problems in obtaining results. While the parameter determination process was carried out in each step, the default values of the other parameters or the values that could be intuitive based on past analysis were used as parameters in the model.

1. Epoch and Batch-Size :

In the first step, parameter selection was performed for appropriate Epoch and Batch-size model parameters. The wider range was first determined for the parameter ranges and then the range and the values in the range were reduced, taking into account the results. The specified parameter ranges are shown in the previous section. Epoch and batch-size values determined as a result of the optimized process applied to each district may vary. The following table shows the parameter values that result from the method applied.

Table 4.10. The epoch and batch-size parameters of Esenyurt and Şişli districts after applying Grid Search method

Esenyurt	Mean Std Test Results
Epochs and Batch Size	0.002126 (0.000682) with: {'batch_size': 60, 'epochs': 50} 0.002062 (0.000457) with: {'batch_size': 60, 'epochs': 100} 0.002086 (0.000490) with: {'batc_size': 60, 'epochs': 200} **0.002012 (0.000539) with: {'batch_size': 60, 'epochs': 500}** 0.002156 (0.000582) with: {'batch_size': 100, 'epochs': 50} 0.002140 (0.000625) with: {'batch_size': 100, 'epochs': 100}

500 - 60	0.002044 (0.000572) with: {'batch_size': 100, 'epochs': 200} 0.002032 (0.000521) with: {'batch_size': 100, 'epochs': 500} 0.002775 (0.000807) with: {'batch_size': 200, 'epochs': 50} 0.002138 (0.000516) with: {'batch_size': 200, 'epochs': 100} 0.002164 (0.000514) with: {'batch_size': 200, 'epochs': 200} 0.002080 (0.000508) with: {'batch_size': 200, 'epochs': 500}		
Şişli	**Mean**	**Std**	**Test Results**
Epochs and Batch Size	0.018861 (0.001038) with: {'batch_size': 60, 'epochs': 100} 0.018996 (0.001113) with: {'batch_size': 60, 'epochs': 200} 0.019534 (0.001072) with: {'batch_size': 60, 'epochs': 500} 0.019003 (0.001070) with: {'batch_size': 100, 'epochs': 100} **0.018830 (0.001030) with: {'batch_size': 100, 'epochs': 200}** 0.018892 (0.001075) with: {'batch_size': 100, 'epochs': 500}		
200-100			

As can be seen from the table, the Epoch and Batch-size values of the districts show different results. The best results in Esenyurt district were 500 Epoch and 60 Batch-size, while in Şişli district 200 Epoch and 100 Batch-size were observed. The model was trained separately for each epoch and batch size using 5-fold cross validation and mean MSE and std values were obtained as in the table above.

2. Activation Function for Input and Hidden Layer

In this step, initially the parameter range included “'relu”, “sigmoid” and “tanh” as well as other functions such as “softmax”, “softsign”, “linear”, and then a reduction in parameter values was made considering the network training results. Considering the results, the most commonly used “relu”, “sigmoid” and “tanh” functions were determined for the parameter list.

3. Activation Function for Output Layer

As in the second step, the same parameter list was determined for the output layer and the appropriate parameters were determined by applying the operation to each district data set. The most suitable parameter was found as "tanh" in all districts. The following table shows the parameter values obtained from Grid Search method for input, hidden and output layers to be used in the analysis to be applied to Bakırköy and Beyoğlu districts.

Table 4.11. The activation function parameters of Bakırköy and Beyoğlu districts after applying Grid Search method

Bakırköy	Mean	Std	Test Results
Activation function (input and hidden layer)	**0.005807 (0.000388) with: {'activation': 'relu'}** 0.006445 (0.000181) with: {'activation': 'tanh'} 0.007708 (0.001566) with: {'activation': 'sigmoid'}		
relu			
Activation function (output layer)	0.819599 (0.002970) with: {'activation': 'relu'} **0.005572 (0.000515) with: {'activation': 'tanh'}** 0.819600 (0.002970) with: {'activation': 'sigmoid'}		
tanh			
Beyoğlu	**Mean**	**Std**	**Test Results**
Table 4.11. (Continued)			
Activation function (input and hidden layer)	0.018201 (0.008480) with: {'activation': 'relu'} 0.018633 (0.010370) with: {'activation': 'tanh'} **0.017098 (0.007025) with: {'activation': 'sigmoid'}**		
sigmoid			
Activation function (output layer)	0.881061 (0.007186) with: {'activation': 'relu'} **0.019668 (0.010180) with: {'activation': 'tanh'}** 0.881622 (0.007260) with: {'activation': 'sigmoid'}		
tanh			

While the best parameter for the output layer was "tanh" in both districts, it was observed that the activation function to be selected for the other layers was different. While the best parameter for input and hidden layers is "relu" in Bakırköy district, this parameter is "sigmoid" in Beyoğlu district. This may be due to the fact that districts' data sets are different or the selected features vary.

4. Dropout Value

In the determination of this parameter, previous analysis were also taken into consideration and the optimization process was performed by using the values in the

specified parameter range. After analysis with this parameter, it was found that there was no result greater than 0.5 and generally the values were in the range of 0.2 to 0.4. Therefore, a parameter list in the range of 0.2 to 0.5 was created for the optimization process.

Table 4.12. The dropout parameters of Silivri, Arnavutköy and Ümraniye districts after applying Grid Search method

Silivri	Mean	Std	Test Results
Dropout	0.004954 (0.005727) with: {'dropout_rate': 0.2}		
	0.004903 (0.005241) with: {'dropout_rate': 0.3}		
0.4	**0.004883 (0.004829) with: {'dropout_rate': 0.4}**		
	0.006119 (0.005197) with: {'dropout_rate': 0.5}		
Arnavutköy	**Mean**	**Std**	**Test Results**
Dropout	**0.012907 (0.001859) with: {'dropout_rate': 0.2}**		
	0.013547 (0.002207) with: {'dropout_rate': 0.3}		
0.2	0.013800 (0.001972) with: {'dropout_rate': 0.4}		
	0.013867 (0.001789) with: {'dropout_rate': 0.5}		

Table 4.12. (Continued)

Ümraniye	Mean	Std	Test Results
Dropout	0.008282 (0.001046) with: {'dropout_rate': 0.2}		
	0.007850 (0.000882) with: {'dropout_rate': 0.3}		
0.3	0.010219 (0.001779) with: {'dropout_rate': 0.4}		
	0.012378 (0.001131) with: {'dropout_rate': 0.5}		

The table above shows the dropout values obtained after applying Grid Search to Silivri, Arnavutköy and Ümraniye districts. According to the results of the analysis with Grid Search, the lowest MSE value in these districts was at different dropout values. The data set size of these districts, the selected features and the data of these districts may be among the factors that make the dropout value differences.

Table 4.13. The hyperparameters of districts after applying Grid Search method

Districts	Epoch – Batch-size	Activition Function (for Input Layer and Hidden Layer)	Activition Function (for Output Layer)	Dropout
Esenyurt	500 - 60	relu	tanh	0.3
Kadıköy	200-60	sigmoid	tanh	0.2
Maltepe	200 -100	relu	tanh	0.2
Beylikdüzü	200-100	sigmoid	tanh	0.2
Ümraniye	500-60	sigmoid	tanh	0.3
Bahçelievler	500-60	sigmoid	tanh	0.2
Başakşehir	500-60	sigmoid	tanh	0.2
Pendik	200-60	sigmoid	tanh	0.2
Eyüp	500-60	sigmoid	tanh	0.2
Kartal	500-60	tanh	tanh	0.2
Avcılar	500-60	relu	tanh	0.2
Ataşehir	500-60	relu	tanh	0.2
Çekmeköy	200-60	tanh	tanh	0.2
Üsküdar	500-60	relu	tanh	0.2
Bağcılar	500-100	relu	tanh	0.2
Table 4.13. (Continued)				
Tuzla	200-60	tanh	tanh	0.2
Kağıthane	500-100	relu	tanh	0.2
Gaziosmanpaşa	500-100	relu	tanh	0.2
Şişli	200-100	sigmoid	tanh	0.3
Bakırköy	500-100	relu	tanh	0.3
Beşiktaş	500-60	relu	tanh	0.2
Büyükçekmece	500-100	relu	tanh	0.2
Sarıyer	500-100	relu	tanh	0.3
Fatih	500-60	tanh	tanh	0.3
Esenler	200-60	relu	tanh	0.2
Güngören	200-60	relu	tanh	0.2
Arnavutköy	500-60	relu	tanh	0.2
Bayrampaşa	500-100	tanh	tanh	0.2
Zeytinburnu	500-60	tanh	tanh	0.2
Silivri	200-60	tanh	tanh	0.4
Sultanbeyli	500-60	tanh	tanh	0.2

Beyoğlu	500-100	sigmoid	tanh	0.2
Beykoz	500-60	tanh	tanh	0.2
Şile	500-60	tanh	tanh	0.2
Çatalca	500-60	relu	tanh	0.2
Adalar	200-100	sigmoid	tanh	0.3

This table shows the parameter values obtained from the Grid Search method applied to all districts' data sets. These selected parameters indicate the similarity or variety of the districts' data sets. In general, it was observed that the dropout value was 0.2 and the epoch value was 500 in 25 districts and 200 in other districts. It can be assumed that the epoch number is not related to the size of the data set, but will vary according to the content of the data. The activation function to be used in the input and hidden layers was found to be "relu" in 16 districts, "sigmoid" in 10 districts and "tanh" in 10 districts. In the output layer, "tanh" function is selected as the best parameter in all districts. Determining the correct parameters to be used in the model with Grid Search analysis has an important role in the success of the model. These parameters were selected as model parameters in the design stage of the ANN model which will be discussed in the next section.

4.4 ANN Design Evaluation

In the last stage, the ANN method was applied to the data sets belonging to the districts. In the ANN models developed to estimate the prices of the residences in districts, the forward propagation algorithm as a learning algorithm was used. In the two hidden layers of ANN model, to determine the most appropriate number of neurons to the problem in the hidden layer, experiments were performed on the number of neurons in hidden layers and different neuron numbers were determined according to the performance results. In the data sets of some districts, neuron numbers were used equally. As a result of the comparison of the estimated values of the model, the most suitable the number of neuron were determined according to RMSE, MAE, MSE and MAPE performance criteria. For other model parameters, the parameters obtained by Grid Search technique mentioned in the previous section were selected. These parameters were discussed in detail and these parameters were evaluated by the ANN model for each district.

MSE is selected as the loss function in the model. The loss function is the function that measures the error rate of the designed model as well as its performance. The loss function is

defined also by the names that are used in the optimization terminology of the objective function or cost function because in the error calculations, it solves the problem by converting the problem into an optimization problem. In the model, the outputs performed with each epoch period are placed in the loss function. The loss function measures the proximity of the model to the correct results according to the results. If the outputs are too many, the value returned by the loss function is high and if the outputs are less, the value of the lost function is low. Then the model updates the weight values of neurons by using selected optimization algorithm with the simplest way to minimize the value of the loss function from the last layers to the first layers in the back propagation models. The loss function differs according to the structure of the model. There is a relationship between the values of normalized education data and loss function. In this study, MSE and MAPE were tried as loss functions. The best performance was seen when MSE was selected as the loss function.

4.4.1 Model Performance Criteria and Results

RMSE, MAPE, MSE, MAE indicators were used in this study for comparing the performance of the models created. (Limsombunchai, 2004: 7).

RMSE is an indicator used to determine the rate of error between the measured values and the values estimated by a model. Therefore, the approximation of the RMSE value to zero indicates that the predictive power of the model has increased. RMSE is calculated by this formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_{pred,i} - y_{actual,i}\right)^2} \quad (3)$$

In the formula, y represents the model, i represents the estimation of the model, y represents the eye, i represents the measured value. n is the number of observations. The MAE is used to show the absolute error between the measured values and the values estimated by the model. Similar to the RMSE value, the approaching of MAE value to zero means the predictive power of the model increases.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|y_{actual,i} - y_{pred,i}\right| \quad (4)$$

The MSE shows the proximity of a regression curve to a set point. The MSE measures estimation performance in a machine learning model and it can be said that near-zero predictions have good performance.

$$MSE = \frac{1}{n}\sum_{j=1}^{n}\left(y_{actual,i} - y_{pred,i}\right)^2 \quad (5)$$

In MAPE regression and time series models, it is frequently used to measure the accuracy of the estimations. If there are zero values between the actual values, MAPE cannot be calculated because there will be a division with zero. The percentage error for very low estimation values cannot exceed 100%, but there is no upper limit for the percentage error when there are very high estimation values.

$$MAPE = 100\frac{\sum_{j=1}^{n}\frac{|y_{actual,i} - y_{pred,i}|}{y_{actual,i}}}{n} \quad (6)$$

4.4.2 k-Fold Cross Validation in ANN Model

In the study, ANN model and k-fold cross-validation were used. Instead of dividing the data sets into "train" and "test" with hold-out management, the data sets were divided into 5 parts using the k-fold cross validation method and it was analyzed with the ANN model. k was chosen as 10, then 5, and it was found that there was no difference in the results when performance evaluation was performed. On the contrary, it was observed that there was an extra time loss when selecting 10 layers. As a result, performances were analyzed in detail by applying 5 fold cross validation and ANN model for each district.

Table 4.14. The performance results of the ANN model with 5 fold cross validation in Kartal district

KARTAL (Normalization [-1,1]) (Time : 50 min)
Score Function : chi-square
(Test 1) 20 Features

Cross-validation Results :

MAE: 0.0131 *MAPE: 1.50%* *RMSE: 0.0260* *MSE: 0.0007*	*MAE: 0.0154* *MAPE: 1.76%* *RMSE: 0.0274* *MSE: 0.0008*	*MAE: 0.0135* *MAPE: 1.66%* *RMSE: 0.0290* *MSE: 0.0008*	*MAE: 0.0130* *MAPE: 1.49%* *RMSE: 0.0271* *MSE: 0.0007*	*MAE: 0.0127* *MAPE: 1.66%* *RMSE: 0.0254* *MSE: 0.0006*

Mean MSE
0.0007

Mean MAE
0.0135

Mean RMSE
0.02698

Mean MAPE
1.61 %

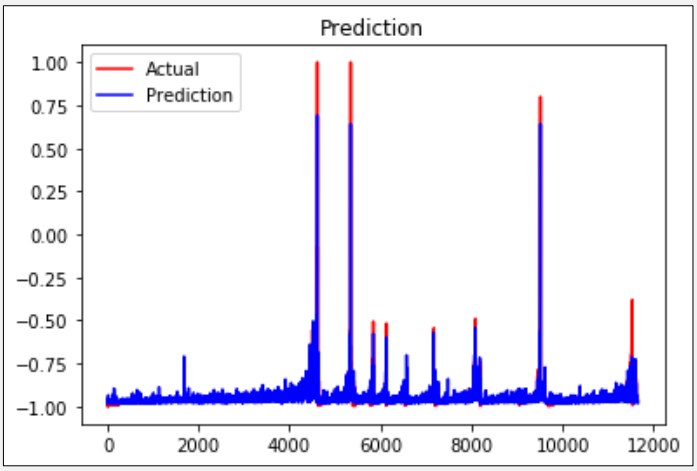


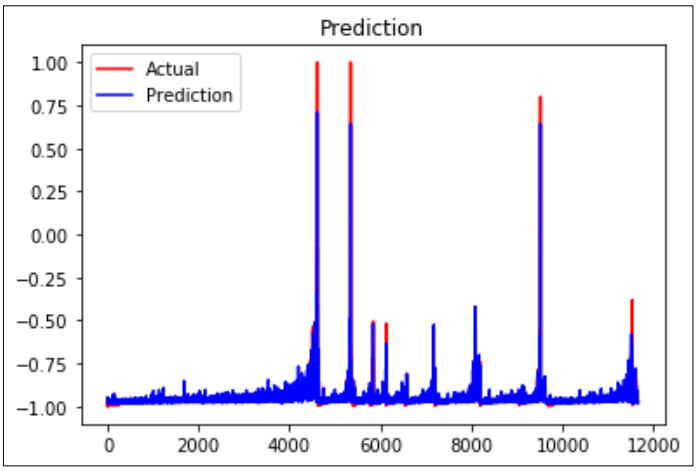
Figure 4.1. Distribution of the actual and the estimation prices - 1

Table 4.14. (Continued)

(Test 2) 50 Features (Time : 55 min)

Cross-validation Results :

MAE: 0.0120 *MAPE: 1.37%* *RMSE: 0.0233* *MSE: 0.0005*	*MAE: 0.0150* *MAPE: 1.79%* *RMSE: 0.0333* *MSE: 0.0011*	*MAE: 0.0124* *MAPE: 1.57%* *RMSE: 0.0248* *MSE: 0.0006*	*MAE: 0.0124* *MAPE: 1.43%* *RMSE: 0.0264* *MSE: 0.0007*	*MAE: 0.0117* *MAPE: 1.60%* *RMSE: 0.0240* *MSE: 0.0006*

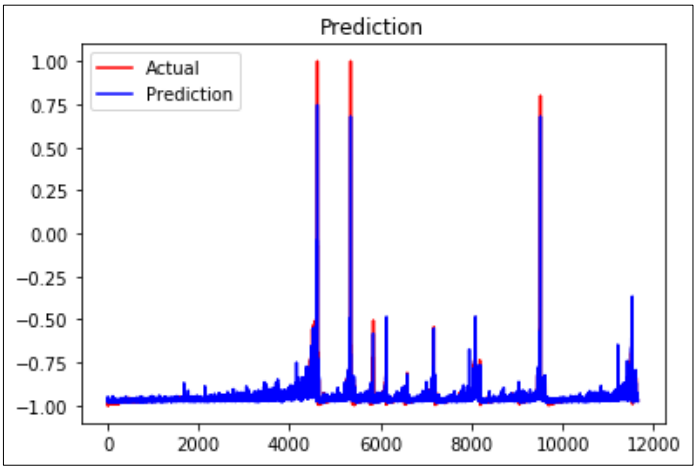
Mean MSE **0.0007** Mean MAE **0.0127** Mean RMSE **0.0270** Mean MAPE **1.55 %**	 **Figure 4.2.** Distribution of the actual and the estimation prices - 2

(Test 3) 100 Features (Time : 1 hour)

Cross-validation Results :

MAE: 0.0114 *MAPE: 1.28%* *RMSE: 0.0211* *MSE: 0.0004*	*MAE: 0.0106* *MAPE: 1.26%* *RMSE: 0.0233* *MSE: 0.0005*	*MAE: 0.0125* *MAPE: 3.34%* *RMSE: 0.0274* *MSE: 0.0008*	*MAE: 0.0108* *MAPE: 1.24%* *RMSE: 0.0242* *MSE: 0.0006*	*MAE: 0.0107* *MAPE: 1.24%* *RMSE: 0.0173* *MSE: 0.0003*

Table 4.14. (Continued)

Mean MSE **0.0005** Mean MAE **0.0112** Mean RMSE **0.02266** Mean MAPE **1.67 %**	 **Figure 4.3.** Distribution of the actual and the estimation prices - 3

Score Function : f-regression

(Test 1) 20 Features (Time : 50 min)

Cross-validation Results :

MAE: 0.0125 *MAPE: 1.39%* *RMSE: 0.0246* *MSE: 0.0006*	*MAE: 0.0143* *MAPE: 1.66%* *RMSE: 0.0280* *MSE: 0.0008*	*MAE: 0.0121* *MAPE: 1.60%* *RMSE: 0.0242* *MSE: 0.0006*	*MAE: 0.0129* *MAPE: 1.48%* *RMSE: 0.0266* *MSE: 0.0007*	*MAE: 0.0118* *MAPE: 1.60%* *RMSE: 0.0235* *MSE: 0.0006*

Mean MSE
0.0006

Mean MAE :
0.0127

Mean RMSE
0.0253

Mean MAPE
1.54 %

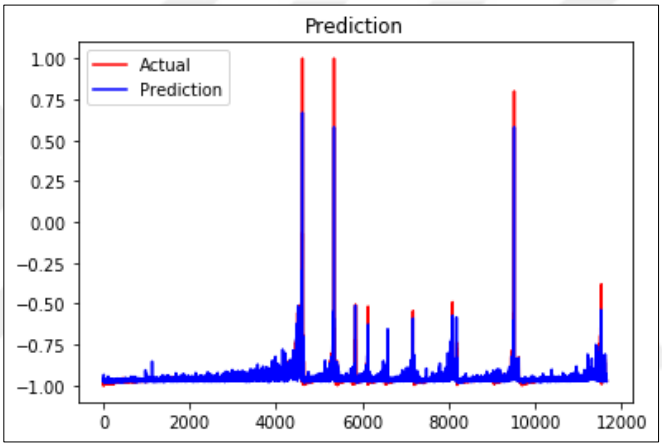


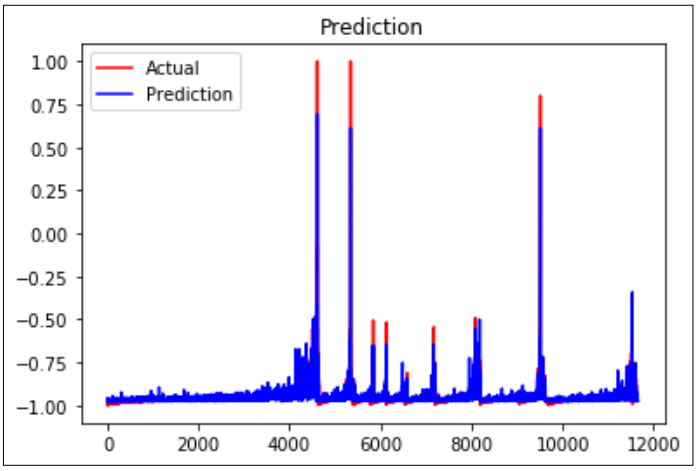
Figure 4.4. Distribution of the actual and the estimation prices - 4

Table 4.14. (Continued)

(Test 2) 50 Features (Time : 55 min)

Cross-validation Results :

MAE: 0.0116 *MAPE: 1.31%* *RMSE: 0.0222* *MSE: 0.0005*	*MAE: 0.0127* *MAPE: 1.42%* *RMSE: 0.0221* *MSE: 0.0005*	*MAE: 0.0118* *MAPE: 2.77%* *RMSE: 0.0250* *MSE: 0.0006*	*MAE: 0.0111* *MAPE: 1.27%* *RMSE: 0.0240* *MSE: 0.0006*	*MAE: 0.0122* *MAPE: 1.64%* *RMSE: 0.0221* *MSE: 0.0005*

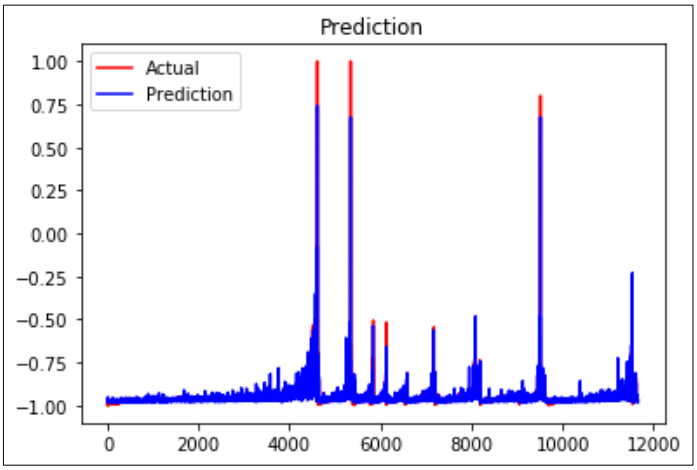
Mean MSE **0.0005** Mean MAE **0.0118** Mean RMSE **0.02308** Mean MAPE **1.68 %**	 **Figure 4.5.** Distribution of the actual and the estimation prices - 5

(Test 3) 100 Features (Time : 1 hour)

Cross-validation Results :

MAE: 0.0106 *MAPE: 1.19%* *RMSE: 0.0209* *MSE: 0.0004*	*MAE: 0.0112* *MAPE: 1.33%* *RMSE: 0.0258* *MSE: 0.0007*	*MAE: 0.0121* *MAPE: 3.45%* *RMSE: 0.0261* *MSE: 0.0007*	*MAE: 0.0106* *MAPE: 1.23%* *RMSE: 0.0241* *MSE: 0.0006*	*MAE: 0.0113* *MAPE: 1.38%* *RMSE: 0.0193* *MSE: 0.0004*

Table 4.14. (Continued)

Mean MSE **0.0005** Mean MAE **0.0113** Mean RMSE **0.0232** Mean MAPE **1.71 %**	 **Figure 4.6.** Distribution of the actual and the estimation prices - 6

The above table shows the performance results of the ANN model with 5 fold cross validation applied in Kartal district. These processes were applied to each district separately as in the table and six different analysis were performed in district data sets. Two different score functions in the SelectKBest method mentioned in the previous section were used. For each score function, three different feature numbers were selected in the districts. In the applied system, the data set was divided into five parts with 5 fold cross validation, four parts were used as training and one part was used as a test. This process was repeated five times in this way and the average performance values were obtained from each analysis. These performance values are MAPE, MSE, MAE and RMSE values as shown in the table and the results for each district will be given in the next table. The performance values of this district (Kartal) were observed to be the best compared to other districts. As can be seen from the forecast graph, it can be said that the actual and estimated prices of this district are very close to each other.

In addition, analysis time was calculated for each district. For some districts (those with large data sets), the analysis time was quite long. It was found to be approximately one hour when the processing time of each analysis was calculated for this district. In the first step after the model is applied, the price estimation, feature selections and mean error information for each fold (5-fold cross validation) are detailed as in the table above. In the last stage, based on this information, summary performance information was created according to district data sets and performances were compared. In the applied model, it is seen that there are different numbers of neurons according to districts. The number of neurons determined in hidden layers is generally larger than the input layer. While selecting the number of neurons in the hidden layers, the number of neurons such as [30, 50, 100, 200] were tested on the model by taking into consideration the selected feature numbers. In some districts, it was observed that performance decreased as the neuron number increased. Therefore, considering the analysis or intuitively, the average neuron numbers were also applied to the model.

The following table shows the neuron numbers and performance results determined in the analysis applied to all district data sets. The performance results in the table is presented in order of order from big to small according to districts data.

Table 4.15. The performance results of the ANN model in the districts' data set (Part 1)

Selection Method (score function of SelectKBest)	Features Number	Hidden Layers Neuron Number	Mean MSE	Mean MAE	Mean RMSE	Mean MAPE (%)
Esenyurt (Size of Data Set : 24.224)						
chi2	20	50 -50	0.0019	0.0245	0.0436	4.20
chi2	50	70-70	0.0017	0.0241	0.0415	4.00
chi2	100	100-100	0.0015	0.0214	0.0391	3.70
f-regression	20	50-50	0.0021	0.0325	0.0456	4.30
f-regression	50	70-70	0.0017	0.0238	0.0410	4.00
f-regression	100	100-100	0.0015	0.0222	0.0394	3.80
Kadıköy (Size of Data Set : 21.579)						
chi2	20	50 -50	0.0033	0.0245	0.0581	6.40
chi2	50	70-70	0.0031	0.0234	0.0558	6.40
chi2	100	100-100	0.0025	0.0206	0.0503	5.90
f-regression	20	50-50	0.0031	0.0236	0.0564	8.03
f-regression	50	70-70	0.0026	0.0220	0.0516	5.90
f-regression	100	100-100	0.0023	0.0202	0.0483	5.30
Maltepe (Size of Data Set : 20.321)						
chi2	20	50 -50	0.0017	0.0235	0.0412	3.00
chi2	50	70-70	0.0012	0.0203	0.0347	2.57
Table 4.15. (Continued)						
chi2	100	100-100	0.0010	0.0179	0.0322	2.26
f-regression	20	50-50	0.0018	0.0240	0.0422	3.00
f-regression	50	70-70	0.0014	0.0217	0.0375	2.70
f-regression	100	100-100	0.0011	0.0192	0.0343	2.45
Beylikdüzü (Size of Data Set : 19.729)						
chi2	20	50 -50	0.0033	0.0338	0.0576	4.77
chi2	50	70-70	0.0025	0.0310	0.0503	4.44
chi2	100	100-100	0.0022	0.0286	0.0469	4.10
f-regression	20	50-50	0.0030	0.0338	0.0551	4.93
f-regression	50	70-70	0.0025	0.0303	0.0500	4.50
f-regression	100	100-100	0.0022	0.0282	0.0469	4.06

Ümraniye (Size of Data Set : 17.407)						
chi2	20	50 -50	0.0012	0.0185	0.0348	2.45
chi2	50	70-70	0.0010	0.0170	0.0318	2.27
chi2	100	100-100	0.0008	0.0156	0.0297	2.08
f-regression	20	50-50	0.0013	0.0192	0.0368	2.59
f-regression	50	70-70	0.0010	0.0171	0.0321	2.24
f-regression	100	100-100	0.0008	0.0152	0.0296	2.03
Bahçelievler (Size of Data Set : 16.488)						
chi2	20	50 -50	0.0017	0.0206	0.0423	2.81
chi2	50	70-70	0.0014	0.0180	0.0381	2.45
chi2	100	100-100	0.0013	0.0164	0.0972	2.31
f-regression	20	50-50	0.0021	0.0221	0.0463	3.10
f-regression	50	70-70	0.0014	0.0189	0.0385	2.59
f-regression	100	100-100	0.0012	0.0164	0.0359	2.27
Başakşehir (Size of Data Set : 16.294)						
chi2	20	50 -50	0.0041	0.0366	0.0643	7.03
chi2	50	70-70	0.0037	0.0345	0.0615	6.55
chi2	100	100-100	0.0034	0.0315	0.0589	6.09
f-regression	20	50-50	0.0051	0.0432	0.0717	7.92
f-regression	50	70-70	0.0041	0.0359	0.0640	6.80
f-regression	100	100-100	0.0036	0.0328	0.0603	6.22
Pendik (Size of Data Set : 16.161)						
chi2	20	50 -50	0.0029	0.0316	0.0541	5.87
Table 4.15. (Continued)						
chi2	50	70-70	0.0027	0.0299	0.0519	4.87
chi2	100	100-100	0.0022	0.0274	0.0468	4.57
f-regression	20	50-50	0.0034	0.0333	0.0585	6.38
f-regression	50	70-70	0.0026	0.0298	0.0514	5.26
f-regression	100	100-100	0.0023	0.0271	0.0477	4.66
Eyüp (Size of Data Set : 13.414)						
chi2	20	50 -50	0.0023	0.0188	0.0473	2.82
chi2	50	70-70	0.0018	0.0162	0.0419	2.36
chi2	100	100-100	0.0016	0.0145	0.0396	2.16
f-regression	20	50-50	0.0028	0.0203	0.0325	2.99
f-regression	50	70-70	0.0017	0.0163	0.0411	2.33
f-regression	100	100-100	0.0015	0.0143	0.0390	2.11

Kartal (Size of Data Set : 11.657)						
chi2	20	50 -50	0.0007	0.0135	0.0270	1.61
chi2	50	70-70	0.0007	0.0127	0.0270	1.55
chi2	100	100-100	0.0005	0.0112	0.0227	1.67
f-regression	20	50-50	0.0005	0.0127	0.0254	1.54
f-regression	50	70-70	0.0005	0.0119	0.0231	1.68
f-regression	100	100-100	0.0005	0.0113	0.0232	1.71
Avcılar (Size of Data Set : 10.367)						
chi2	20	50 -50	0.0041	0.0337	0.0633	5.73
chi2	50	70-70	0.0041	0.0317	0.0635	5.58
chi2	100	100-100	0.0041	0.0292	0.0636	5.32
f-regression	20	50-50	0.0050	0.0378	0.0701	6.28
f-regression	50	70-70	0.0047	0.0342	0.0680	5.82
f-regression	100	100-100	0.0045	0.0318	0.0662	5.59

The analysis results of the districts of Esenyurt, Kadıköy, Maltepe, Beylikdüzü, Ümraniye, Bahçelievler, Başakşehir, Pendik, Eyüp, Kartal and Avılar, which contain between 10.000 and 25.000 data, are shown in the above table. When the results of Esenyurt district with the largest number of data are examined, it is observed that there is no significant difference when the SelectKBest feature selection methods (according to score functions) applied are compared. As a result of the analysis performed by selecting 20 features, 50 features and 100 features, it was observed that MSE, MAE, RMSE and MAPE values decreased as the number of features and neurons increases. This is the case for both feature selections made with the chi2 and f_regression score function. When the districts in the table were compared, the lowest MSE value of 0.0015 was found in the analysis with 100 feature selection and hidden layers with 100 neurons. The lowest MAPE value of 3.70% was obtained by model analysis using SelectKBest (score function: chi2) method and having 100 neurons in input and hidden layers. It is observed that other performance values are also low values as a result of this analysis. Kadikoy district also has similar results to Esenyurt district. In contrast, the lowest MSE (0.0023), MAE (0.0202), RMSE (0.0483) and MAPE (5.30%) values in this district were obtained with the following parameters; input layer with 100 neurons (SelectKBest method with score function f_regression was used) and hidden layers with 100 neurons. The results of the analysis in Maltepe district are similar to those of

Esenyurt district. In Beylikdüzü, Ümraniye, Bahçelievler and Eyüp districts, the best performance values were reached with the following parameters; input layer with 100 neurons (SelectKBest method with score function f_regression was used) and hidden layers with 100 neurons. It can be concluded that the performance values may be related to the data content rather than the size of the data set when the districts' analysis in the table above are compared. Kartal district with approximately 11.000 data has the lowest MSE (0.0005), MAE (0.0112), RMSE (0.0227) and MAPE (1.54) values. It is seen that this district is one of the districts with the least data according to the table. At the same time, the analysis results of this district are found to be better than the results of the districts' analysis in the following tables.

Table 4.16. The performance results of the ANN model in the districts' data set (Part 2)

Selection Method (score function of SelectKBest)	Features Number	Hidden Layers Neuron Number	Mean MSE	Mean MAE	Mean RMSE	Mean MAPE (%)
Ataşehir (Size of Data Set : 9.440)						
chi2	20	50 -50	0.0042	0.0353	0.0638	7.76
chi2	50	70-70	0.0045	0.0324	0.0662	7.61
chi2	30	50 -50	0.0043	0.0346	0.0650	7.56
f-regression	20	50-50	0.0043	0.0358	0.0654	8.53
f-regression	50	70-70	0.0045	0.0343	0.0670	7.70
f-regression	30	50 -50	0.0041	0.0335	0.0639	8.09
Table 4.16. (Continued)						
Çekmeköy (Size of Data Set : 9.221)						
chi2	20	50 -50	0.0014	0.0184	0.0379	2.44
chi2	50	70-70	0.0012	0.0167	0.0355	2.23
chi2	100	100-100	0.0012	0.0172	0.0350	2.28
f-regression	20	50-50	0.0017	0.0540	0.0413	2.69
f-regression	50	70-70	0.0011	0.0177	0.0330	2.31
f-regression	100	100-100	0.0012	0.0164	0.0346	2.23
Üsküdar (Size of Data Set : 8.395)						
chi2	20	50 -50	0.0076	0.0423	0.0720	11.82
chi2	50	70-70	0.0060	0.0279	0.0776	7.95
chi2	100	100-100	0.0068	0.0278	0.0822	8.14

f-regression	20	50-50	0.0082	0.0362	0.0905	11.51
f-regression	50	70-70	0.0072	0.0836	0.0846	9.24
f-regression	100	100-100	0.0065	0.0290	0.0804	8.20
Bağcılar (Size of Data Set : 8.141)						
chi2	20	50 -50	0.0025	0.0243	0.0489	3.27
chi2	50	70-70	0.0022	0.0214	0.0455	2.95
chi2	100	100-100	0.0022	0.0208	0.0457	2.89
f-regression	20	50-50	0.0024	0.0249	0.0477	3.35
f-regression	50	70-70	0.0022	0.0220	0.0448	3.14
f-regression	100	100-100	0.0022	0.0206	0.0454	2.84
Tuzla (Size of Data Set : 6.985)						
chi2	20	50 -50	0.0017	0.0237	0.0418	3.01
chi2	50	70-70	0.0018	0.0228	0.0421	2.87
chi2	100	100-100	0.0016	0.0204	0.0396	2.58
f-regression	20	50-50	0.0020	0.0242	0.0451	3.07
f-regression	50	70-70	0.0016	0.0213	0.0401	2.70
f-regression	100	100-100	0.0015	0.0205	0.0385	2.58
Kağıthane (Size of Data Set : 6.807)						
chi2	20	50 -50	0.0037	0.0372	0.0605	5.99
chi2	50	70-70	0.0030	0.0330	0.0547	5.82
chi2	100	100-100	0.0031	0.0310	0.0562	5.63
f-regression	20	50-50	0.0035	0.0376	0.0592	5.94
f-regression	50	70-70	0.0034	0.0343	0.0585	5.57
Table 4.16. (Continued)						
f-regression	100	100-100	0.0028	0.0310	0.0529	4.80
Gaziosmanpaşa (Size of Data Set : 6.515)						
chi2	20	50 -50	0.0073	0.0565	0.0855	20.09
chi2	50	70-70	0.0068	0.0542	0.0829	18.84
chi2	100	100-100	0.0059	0.0510	0.0771	16.39
f-regression	20	50-50	0.0071	0.0584	0.0846	18.64
f-regression	50	70-70	0.0065	0.0555	0.0810	17.40
f-regression	100	100-100	0.0058	0.0503	0.0763	16.04
Şişli (Size of Data Set : 5.829)						
chi2	20	50 -50	0.0187	0.0918	0.1368	16.11

chi2	50	70-70	0.0187	0.0909	0.1368	16.01
chi2	100	100-100	0.0195	0.0922	0.1397	16.06
f-regression	20	50-50	0.0187	0.0915	0.1366	16.06
f-regression	50	70-70	0.0187	0.0906	0.1366	15.99
f-regression	100	100-100	0.0196	0.0914	0.1401	16.07
Bakırköy (Size of Data Set : 5.697)						
chi2	20	50 -50	0.0036	0.0339	0.0598	19.85
chi2	50	70-70	0.0034	0.0309	0.0585	19.84
chi2	100	100-100	0.0032	0.0287	0.0569	20.62
f-regression	20	50-50	0.0039	0.0347	0.0628	19.72
f-regression	50	70-70	0.0032	0.0310	0.0571	17.95
f-regression	100	100-100	0.0034	0.0308	0.0581	18.83
Beşiktaş (Size of Data Set : 5.421)						
chi2	20	50 -50	0.0037	0.0591	0.0610	5.69
chi2	50	70-70	0.0040	0.0266	0.0635	5.67
chi2	100	100-100	0.0039	0.0281	0.0623	5.54
f-regression	20	50-50	0.0041	0.0318	0.0643	6.18
f-regression	50	70-70	0.0045	0.0289	0.0673	5.72
f-regression	100	100-100	0.0042	0.0297	0.0648	5.94
Büyükçekmece (Size of Data Set : 5.153)						
chi2	20	50 -50	0.0058	0.0348	0.0752	9.33
chi2	50	70-70	0.0057	0.0304	0.0744	8.81
chi2	100	100-100	0.0056	0.0292	0.0740	8.67
f-regression	20	50-50	0.0062	0.0354	0.0775	9.14
Table 4.16. (Continued)						
f-regression	50	70-70	0.0061	0.0932	0.0775	9.11
f-regression	100	100-100	0.0060	0.0312	0.0763	8.93

The analysis results of districts with data between 5.000 and 9.000 are shown in this table. These districts are Ataşehir, Çekmekoy, Üsküdar, Bağcılar, Tuzla, Kağıthane, Gaziosmanpaşa, Şişli, Bakırkoy, Beşiktaş, Büyükçekmece. When the districts in the table are examined, it is seen that the district with the best performance value is Çekmeköy. The best performance values in this district are obtained in different analysis. In this district, the analysis results that finds the best MSE and RMSE values: 50 features selected with SelectKBest (score function: f_regression), Hidden Layers = 70 neurons, the analysis results

that finds the best MAE and MAPE values: selected with SelectKBest (score function: f_regression) 100 features, Hidden Layers = 100 neurons. When the analysis results of Tuzla and Bağcılar districts are examined, it has better performance values than the other districts in the table. The results of Gaziosmanpaşa, Bakırköy and Şişli districts' analysis contain the highest error values. As a result of the analysis carried out in Şişli district, it was observed that the MSE (0.0196) value was the highest. The highest MAPE value (20.62%) was obtained in the analysis of Bakırköy district, in this table. The parameters of this analysis are; 100 features selected with SelectKBest (score function: chi-square) and Hidden Layers = 100 neurons.

Table 4.17. The performance results of the ANN model in the districts' data set (Part 3)

Selection Method (score function of SelectKBest)	**Features Number**	**Hidden Layers Neuron Number**	**Mean MSE**	**Mean MAE**	**Mean RMSE**	**Mean MAPE (%)**
Sarıyer (Size of Data Set : 4.589)						
chi2	20	50 -50	0.0069	0.0446	0.0832	35.38
chi2	50	70-70	0.0070	0.0396	0.0835	32.32
chi2	30	50 -50	0.0063	0.0413	0.0413	22.21
f-regression	20	50-50	0.0086	0.0485	0.0928	42.76
f-regression	50	70-70	0.0075	0.0419	0.0861	44.26
f-regression	30	50 -50	0.0077	0.0441	0.0876	42.00
Table 4.17. (Continued)						
Fatih (Size of Data Set : 4.217)						
chi2	20	50 -50	0.0086	0.0540	0.0925	14.81
chi2	50	70-70	0.0081	0.0524	0.0899	14.49
chi2	100	100-100	0.0071	0.0487	0.0837	12.64
f-regression	20	50-50	0.0087	0.0560	0.0933	15.47
f-regression	50	70-70	0.0077	0.0510	0.0870	14.50
f-regression	100	100-100	0.0076	0.0494	0.0866	13.19
Esenler (Size of Data Set : 4.147)						
chi2	20	50 -50	0.0107	0.0564	0.1025	25.57
chi2	50	70-70	0.0096	0.0536	0.0974	23.42

chi2	100	100-100	0.0090	0.0518	0.0942	25.60
f-regression	20	50-50	0.0123	0.0648	0.1106	23.93
f-regression	50	70-70	0.0104	0.0553	0.0957	24.15
f-regression	100	100-100	0.0094	0.0509	0.0960	25.97
Güngören (Size of Data Set : 3.940)						
chi2	20	50 -50	0.0052	0.0385	0.0717	10.26
chi2	50	70-70	0.0050	0.0370	0.0705	9.88
chi2	100	100-100	0.0052	0.0355	0.0713	8.87
f-regression	20	50-50	0.0056	0.0422	0.0737	11.52
f-regression	50	70-70	0.0055	0.0400	0.0736	10.04
f-regression	100	100-100	0.0048	0.0359	0.0689	8.22
Arnavutköy (Size of Data Set : 3.598)						
chi2	20	50 -50	0.0103	0.0627	0.1013	17.77
chi2	50	70-70	0.0091	0.0588	0.0952	14.98
chi2	30	50 -50	0.0094	0.0607	0.0969	15.77
f-regression	20	50-50	0.0110	0.0681	0.1049	16.36
f-regression	50	70-70	0.0096	0.0595	0.0979	14.36
f-regression	70	100-100	0.0094	0.0587	0.0971	13.77
Bayrampaşa (Size of Data Set : 3.309)						
chi2	20	50 -50	0.0047	0.0428	0.0681	7.03
chi2	50	70-70	0.0047	0.0408	0.0680	6.71
chi2	100	100-100	0.0044	0.0374	0.0657	6.46
f-regression	20	50-50	0.0052	0.0462	0.0714	8.76
f-regression	50	70-70	0.0047	0.0406	0.0683	6.96
Table 4.17. (Continued)						
f-regression	70	100-100	0.0051	0.0381	0.0697	6.98
Zeytinburnu (Size of Data Set : 3.200)						
chi2	20	50 -50	0.0019	0.0217	0.0430	3.06
chi2	50	70-70	0.0017	0.0200	0.0417	2.98
chi2	30	50-50	0.0014	0.0193	0.0375	2.87
f-regression	20	50-50	0.0016	0.0232	0.0405	3.15
f-regression	50	70-70	0.0019	0.0206	0.0428	2.97
f-regression	30	50-50	0.0014	0.0203	0.0376	2.96
Silivri (Size of Data Set : 2.715)						

chi2	20	50 -50	0.0039	0.0288	0.0565	7.12
chi2	50	70-70	0.0036	0.0289	0.0539	5.21
chi2	70	100-100	0.0035	0.0285	0.0525	4.26
f-regression	20	50-50	0.0049	0.0297	0.0612	11.75
f-regression	50	70-70	0.0045	0.0308	0.0597	5.50
f-regression	70	100-100	0.0042	0.0283	0.0551	4.46
Sultanbeyli (Size of Data Set : 2.561)						
chi2	20	50 -50	0.0043	0.0398	0.0652	10.67
chi2	50	70-70	0.0035	0.0360	0.0592	7.09
chi2	70	100-100	0.0034	0.0352	0.0584	7.19
f-regression	20	50-50	0.0046	0.0436	0.0680	11.68
f-regression	50	70-70	0.0040	0.0400	0.0628	7.66
f-regression	70	100-100	0.0035	0.0351	0.0590	6.80
Beyoğlu (Size of Data Set : 2.440)						
chi2	20	50 -50	0.0130	0.0392	0.1020	8.08
chi2	50	70-70	0.0128	0.0376	0.1099	7.40
chi2	70	100-100	0.0097	0.0333	0.0960	6.52
f-regression	20	50-50	0.0130	0.0378	0.0378	7.61
f-regression	50	70-70	0.0124	0.0360	0.1098	7.35
f-regression	70	100-100	0.0110	0.0329	0.1039	6.13
Beykoz (Size of Data Set : 818)						
chi2	20	50 -50	0.0108	0.0820	0.0999	5.91
chi2	50	70-70	0.0085	0.0402	0.0913	5.80
chi2	100	100-100	0.0076	0.0407	0.0857	5.98
Table 4.17. (Continued)						
f-regression	20	50-50	0.0098	0.0410	0.0977	6.33
f-regression	50	70-70	0.0090	0.0406	0.0934	5.82
f-regression	30	50-50	0.0076	0.0380	0.0854	5.36
Şile (Size of Data Set : 517)						
chi2	20	50 -50	0.0127	0.0872	0.1124	22.18
chi2	50	70-70	0.0120	0.0858	0.1094	22.59
chi2	30	50-50	0.0122	0.0859	0.1104	21.77
f-regression	20	50-50	0.0118	0.0856	0.1087	21.83
f-regression	50	70-70	0.0105	0.0769	0.1020	20.06
f-regression	30	50-50	0.0106	0.0791	0.1027	20.13

Çatalca (Size of Data Set : 480)						
chi2	20	50 -50	0.0083	0.0668	0.0905	29.08
chi2	50	70-70	0.0087	0.0643	0.0931	33.07
chi2	30	50-50	0.0081	0.0631	0.0899	32.64
f-regression	20	50-50	0.0112	0.0725	0.1050	36.28
f-regression	50	70-70	0.0092	0.0631	0.0959	35.12
f-regression	100	100-100	0.0092	0.0607	0.0958	33.09
Adalar (Size of Data Set : 274)						
chi2	20	50 -50	0.0180	0.0436	0.1027	4.89
chi2	50	70-70	0.0188	0.0487	0.1081	5.45
chi2	10	30-30	0.0173	0.0408	0.0977	4.55
f-regression	20	50-50	0.0181	0.0441	0.1032	4.93
f-regression	50	70-70	0.0187	0.0491	0.1080	5.49
f-regression	10	30-30	0.0174	0.0418	0.0990	4.65

When the districts' analysis in the third table is examined, the smallest error values are found in the analysis made in Zeytinburnu district. The districts in this table are Sarıyer, Fatih, Esenler, Güngören, Arnavutköy, Bayrampaşa, Zeytinburnu, Silivri, Sultanbeyli, Beyoğlu, Beykoz, Şile, Çatalca and Adalar. The smallest MSE (0.0014) and MAPE (2.87) values according to the districts in the table are observed as the result of analysis with 30 neurons input layer and 50 neurons hidden layers parameters in Zeytinburnu district. Smaller MAE and RMSE values are obtained in the analysis applied in this district compared to the other districts in the table. According to the results in the table, a smaller MAPE value is obtained in Adalar, which the smallest data set, compared to districts containing larger data sets. In addition, the highest MSE value is observed in this district. In the table, the large MAPE, MSE, MAE and RMSE values are noteworthy. It can be assumed that the size of the data set and the content of the data set may have an effect on higher values. When the analysis results in the table were compared, the MAPE value of approximately 40% was obtained in the analysis results in Sarıyer district. According to the analysis applied to Çatalca (approximately 33%) and Şile (approximately 22%), high MAPE values are reached. In addition, high MSE (0.0127) value was reached in Şile district. Again, the highest MAE (0.0872) and RMSE (0.1124) values were determined in this district.

It was observed that a large number of selected features adversely affected the analysis results in some districts. Therefore, a smaller number of features were selected in these districts and the analysis were conducted with the determined parameters. In the Adalar district, three different analysis were performed with feature selection in the form of "20", "50" and "10" and analysis result with "10" feature selection were found to be better and have a smaller error value. When the results of the analysis in Zeytinburnu district were compared, it was observed that the analysis with "30" features had better MSE (0.0014), MAE (0.0193), RMSE (0.0375) and MAPE (2.87%) values.

It was aimed to reach the best performance and estimation values, in the analysis applied to the districts shown in all three tables. With SelectKBest described in the previous section, the best features were tried to be selected, then parameters were optimized with Grid Search technique and the model parameter was determined and 5-fold cross validation technique was applied with ANN and Grid Search methods. After applying these methods step by step, the performances were examined and the results were compared for each district as in the tables above. As a result of the comparison of district analysis results, it can be said that the appropriate data set will provide better performance for the model. Of course, the size of the data set is important for ANN, but as a result of analysis conducted in 36 districts, it is clear that not only the size of the data set, but also the content of the data set is important. Besides, model parameters determined by using the Grid Search method have an important role in the success of the model. Although the application of this technique leads to prolongation of the analysis times, with this technique, more accurate results were obtained with more suitable parameters. The most important advantage of the 5-fold cross validation technique is that the overfitting is avoided in the learning stage of the model and it is intended to achieve real performance values with this technique. The accuracy and performance of the results obtained from the hybrid model developed with these methods are among the most important factors for this study. Therefore, the application phase of these methods has been done carefully and the results have been scrupulously examined.

5. CONCLUSIONS

Real estate valuation is a serious topic which requires the consideration of several complex parameters. In estimating the real estate value, a large number of features are often used to determine the fair value of a property. For real estate price estimation, the traditional approaches are less accurate, because, when making a prediction, it is difficult to identify a number of variables and calculate their effect. The importance of real estate appraisal increases day by day and the scientific studies in this field continue to increase. When the studies on the subject are examined in the literature, many methods have been developed by

analyzing many parameters, determining factors on the value of the real estate and revealing the best combinations accordingly.

Istanbul, which forms the data set of the thesis study, is a more complex city in terms of real estate appraisal compared to other cities and the data used in this study contains many parameters. In addition, choosing a province like Istanbul, the fact that this province has sufficient parameters and coverage makes a significant contribution to achieve successful results by using the ANN method. The most important advantage of the ANN method in real estate appraisal is the ability of the method to solve complex problems. In order to make an accurate appraisal, it is necessary to review all the parameters affecting the real estate price and to digitize the data set used for artificial neural networks. It is important to reach the data that contains the required parameters and to design the used parameters correctly to achieve a highly consistent result with artificial neural networks. Of course, it is very difficult to talk about certain values in real estate valuation. However, it can be said that a more realistic approach has been created through different methods and feature selection. With the feature selection method used in the study, the selection of the factors that affect the real estate price has been facilitated and this method has contributed to the success of the model. By using SelectKBest method for feature selection, different score functions of this method have been tried and the most important features affecting the price have been determined. It is also important to note that the data preprocessing stage is very important for the success of the model. If this stage is not done correctly, it is quite low to produce realistic results from the estimation model. In this study, data preprocessing was applied step by step for all district data sets and the data were made suitable for the model. Although this process is applied, it is determined that there are erroneous and contradictory data that adversely affect the results of the analysis in some of the data sets. Therefore, preprocessing steps were applied for the second time in these districts.

Considering the negative situations of traditional methods, it is important to develop an approach towards real estate valuation through stochastic methods realized with the help of computer technologies for this study. Thanks to the Grid Search method in the research, the model parameters has been optimized and it contributed to the success of this model. ANN parameters obtained with this method were determined before the analysis and it has made this process more reliable.

In addition, the k-fold cross validation technique, which was applied with both Grid Search and ANN methods, contributed significantly to the accuracy of the results. With this technique, the data were divided into 5 fold, the model was trained for each layer and then the model performances were measured. The success of the model was evaluated with the obtained average performance values in this method. Besides, it was aimed to prevent overfitting and to reach accurate results thanks to this technique used in the study.

When the results of the analysis were examined, it was seen that MAPE, RMSE, MSE and MAE values were generally quite low. The data set of each district was compiled separately by the model, and the estimation results of the model were compared and it was observed that the model produced successful estimates. Thanks to the large data set of the majority of the district and the correct data preprocessing steps, it has been an advantage in the performance of the model. Based on the results obtained with the multi-layer perceptron method applied in the model, it can be used in real estate appraisal of artificial neural networks and has been found to be a successful method.

6. RECOMMENDATIONS

The housing market is an important element for the economy. It can be said that many criteria affect housing prices and these criteria vary by region. External factors, especially economic developments, also affect housing prices closely and make it difficult to estimate them. Since the houses are used for shelter as well as for investment purposes, for those who want to invest, it is important that the house price estimations are close to reality and that the correct methods are determined.

The use of classical and traditional methods in the real estate appraisal process may cause the estimated value to differ from the actual market value. The heterogeneous nature of this sector and the need to consider many dependent and independent variables indicate that estimation methods require metaheuristic methods rather than traditional methods. Therefore, with the usage metaheuristic methods. The predictions will be more and more accurate. Although the use of metaheuristic methods has increased in recent years, it is clear that further research will improve the results positively. The artificial neural network model used in this study is quite capable of real estate valuation. As a result of the analysis conducted in almost all districts of Istanbul, very successful estimations have been obtained. As a result of the studies to be made in other provinces by artificial neural networks method, it is predicted that successful results can be obtained in the same way. In this thesis, it is seen that the performance is generally good in the analysis made using the different numbers of selected features. It can be argued that the estimation success may be similar since the house characteristics of different provinces may be similar. Performing more analysis using this method may show a better measurement of the success of the method. The application of this method in different provinces across Turkey may lead to the expansion and the development of the analysis. Furthermore, during a real estate appraisal, the inclusion of different provinces in the data set allows price estimations to be compared by province and this makes it possible to test the differences between the provinces.

Another important point is the accurate data preprocessing, except for the method used to estimate the real estate price. If the data preprocessing is not done correctly, the success rate of the method decreases considerably. That is, the fact that data preprocessing is correct and appropriate to the method to be used affects the success and usability of the method positively. At the same time, although the method is good, incorrect data preprocessing can create a misleading impression about the method used. For this reason, it is clear that taking into account the features of the house, as well as the correct extraction of the data and careful preprocessing of the data set, will increase the success of the methods to be used and will benefit the next studies. For real estate price estimation, all stages should be listed in an appropriate manner in the analysis to be performed and each stage should be performed carefully. In addition, the feature selection process, which is an important part of these stages, has a significant impact on the estimated value of the price. The proper determination of the features of the house also affects the success of the forecast positively. Briefly, all the

parameters required for real estate price estimation should be considered and calculated correctly. It should be taken into consideration that if the methods are improved the results will be more and more effective.

REFERENCES

Ahn, J. J., Byun, H. W., Oh, K. J., Kim, T. Y. (2012). Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. Expert Systems with Applications, 39(9), 8369-8379.

Bagnoli, C., & Smith, H. (1998). The theory of fuzz logic and its application to real estate valuation. Journal of Real Estate Research, 16(2), 169-200.

Bee-Hua, G. (2000). Evaluating the performance of combining neural networks and genetic algorithms to forecast construction demand: the case of the Singapore residential sector. Construction Management & Economics, 18(2), 209-217.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13(Feb), 281-305.

Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. Critical care, 7(6), 451.

Borst, R. A. (1991). Artificial neural networks: the next modelling/calibration technology for the assessment community. Property Tax Journal, 10(1), 69-94.

Broomhead, D. S., & Lowe, D. (1988). Radial basis functions, multi-variable functional interpolation and adaptive networks (No. RSRE-MEMO-4148). Royal Signals and Radar Establishment Malvern (United Kingdom).

Byrne, P. (1995). Fuzzy analysis: a vague way of dealing with uncertainty in real estate analysis?. Journal of Property Valuation and Investment, 13(3), 22-41.

Bulsari, A. B., & Saxén, H. (1995). A recurrent network for modeling noisy temporal sequences. Neurocomputing, 7(1), 29-40.

Calhoun, C. A. (2003). Property valuation models and house price indexes for the provinces of Thailand: 1992-2000. Housing Finance International, 17(3), 31-41.

Case, K. E., & Shiller, R. J. (1987). Prices of single family homes since 1970: New indexes for four cities.

Case, K. E., & Shiller, R. J. (1988). The efficiency of the market for single-family homes.

Case, K. E., Shiller, R. J. (1990). Forecasting prices and excess returns in the housing market. Real Estate Economics, 18(3), 253-273.

Chiarazzo, V., Caggiani, L., Marinelli, M., Ottomanelli, M. (2014). A Neural Network based model for real estate price estimation considering environmental quality of property location. Transportation Research Procedia, 3, 810-817.

Coskun, Y. (2011). The establishment of the real estate regulation and supervision agency of Turkey (RERSAT). Housing Finance International, 25(4), 42-51.

Das, S., Gupta, R., & Kabundi, A. (2009). Could we have predicted the recent downturn in the South African housing market?. Journal of Housing Economics, 18(4), 325-335.

Del Giudice, V., De Paola, P., & Forte, F. (2017). Using genetic algorithms for real estate appraisals. Buildings, 7(2), 31.

Del Giudice, V., De Paola, P., & Cantisani, G. B. (2017). Valuation of real estate investments through Fuzzy Logic. Buildings, 7(1), 26.

Do, A. Q., & Grudnitski, G. (1992). A neural network approach to residential property appraisal. The Real Estate Appraiser, 58(3), 38-45.

Du, D., Li, A., Zhang, L. (2014). Survey on the applications of big data in Chinese real estate enterprise. Procedia Computer Science, 30, 24-33.

Du, D., Li, A., Zhang, L., Li, H. (2014). Review on the applications and the handling techniques of big data in Chinese realty enterprises. Annals of Data Science, 1(3-4), 339-357.

Fausett, L. V. (1994). Fundamentals of neural networks: architectures, algorithms, and applications (Vol. 3). Englewood Cliffs: prentice-Hall.

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. Journal of machine learning research, 3(Mar), 1289-1305.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model: one-sided estimation and forecasting. Journal of the American Statistical Association, 100(471), 830-840.

Ge, J. X., Runeson, G., & Lam, K. C. (2003). Forecasting Hong Kong housing prices: An artificial neural network approach. In International conference on methodologies in housing research, Stockholm, Sweden.

Geisser, S. (1975). The predictive sample reuse method with applications. Journal of the American statistical Association, 70(350), 320-328.

Gerek, I. H. (2014). House selling price assessment using two different adaptive neuro-fuzzy techniques. Automation in Construction, 41, 33-39.

Gogas, P., Pragidis, I. (2013). Does the interest risk premium predict housing prices. DUTH Research Papers in Economics 1-2013, Democritus University of Thrace, Department of Economics.

Gu, J., Zhu, M., & Jiang, L. (2011). Housing price forecasting based on genetic algorithm and support vector machine. Expert Systems with Applications, 38(4), 3383-3386.

Gupta, R. (2013). Forecasting house prices for the four census regions and the aggregate US economy in a data-rich environment. Applied Economics, 45(33), 4677-4697.

Gupta, R., Das, S. (2010). Predicting downturns in the US housing market: A Bayesian approach. The Journal of Real Estate Finance and Economics, 41(3), 294-319.

Gupta, R., Kabundi, A. (2010). Forecasting real US house prices: principal components versus bayesian regressions. International Business & Economics Research Journal (IBER), 9(7), 141-152.

Guan, J. and A. Levitan. Artificial Neural Network-Based Assessment of Residential Real Properties: A Case Study. Accounting Forum, 1996, 20:3–4, 311–26.

Göncü, İ., & Stratejileri, K. P. (2004). İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü (Doctoral dissertation, Yüksek Lisans Tezi, İstanbul).

Haykin, S. (1994). Neural networks: a comprehensive foundation. Prentice Hall PTR.

Hasiloglu, A., Yilmaz, M., Comakli, O., & Ekmekci, I. (2004). Adaptive neuro-fuzzy modeling of transient heat transfer in circular duct air flow. International journal of thermal sciences, 43(11), 1075-1090.

Hromada, E. (2015). Mapping of real estate prices using data mining techniques. Procedia Engineering, 123, 233-240.

Khamis, A. B., & Kamarudin, N. K. K. B. (2014). Comparative Study On Estimate House Price Using Statistical And Neural Network Model. International Journal of Scientific & Technology Research, 3(12), 126-131.

Khalafallah, A. (2008). Neural network based model for predicting housing market performance. Tsinghua Science and Technology, 13(S1), 325-328.

Kauko, T., Hooimeijer, P., & Hakfoort, J. (2002). Capturing housing market segmentation: An alternative approach based on neural network modelling. Housing Studies, 17(6), 875-894.

Kettani, O., Oral, M. (2015). Designing and implementing a real estate appraisal system: The case of Québec Province, Canada. Socio-Economic Planning Sciences, 49, 1-9.

Kocer, H. E., & Albayrak, U. (2015). Measuring The Effect Of Multi-Touch Panel Based Education For Pre-School Students. Procedia-Social and Behavioral Sciences, 191, 1560-1570.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling(Vol. 26). New York: Springer

Kummerow, M. (2007) Price differences models, 13th Pacific-Rim Real Estate Society Conference,Fremantle, Western Australia, January.

Kuşan, H., Aytekin, O., & Özdemir, İ. (2010). The use of fuzzy logic in predicting house selling price. Expert systems with Applications, 37(3), 1808-1813.

Küçükaslan, N. (2015). Emlak Pazarlaması. Ekin Basım Yay., Bursa.

Lam, K. C., Yu, C. Y., & Lam, K. Y. (2008). An artificial neural network and entropy model for residential property price forecasting in Hong Kong. Journal of Property Research, 25(4), 321-342.

Ladha, L., & Deepa, T. (2011). Feature selection methods and algorithms. International journal on computer science and engineering, 3(5), 1787-1797.

Lenk, M. M., Worzala, E. M., & Silva, A. (1997). High-tech valuation: should artificial neural networks bypass the human valuer?. Journal of Property Valuation and Investment, 15(1), 8-26.

Limsombunchai, V. (2004, June). House price prediction: hedonic price model vs. artificial neural network. In New Zealand Agricultural and Resource Economics Society Conference (pp. 25-26).

Linneman, P. (1986). An empirical test of the efficiency of the housing market. Journal of Urban Economics, 20(2), 140-154.

Manganelli, B., De Mare, G., & Nesticò, A. (2015, June). Using genetic algorithms in the housing market analysis. In International Conference on Computational Science and Its Applications (pp. 36-45). Springer, Cham.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4), 115-133.

McGreal, S., Adair, A., McBurney, D., & Patterson, D. (1998). Neural networks: the prediction of residential values. Journal of Property Valuation and Investment, 16(1), 57-70.

McLeod, S. (2007). Maslow's hierarchy of needs. Simply psychology, 1.

Mehra, P., & Wah, B. W. (1992, June). Adaptive load-balancing strategies for distributed systems. In Proceedings of the Second International Conference on Systems Integration (pp. 666-675). IEEE.

Mukhlishin, M. F., Saputra, R., & Wibowo, A. (2017, November). Predicting house sale price using fuzzy logic, Artificial Neural Network and K-Nearest Neighbor. In Informatics and Computational Sciences (ICICoS), 2017 1st International Conference on (pp. 171-176). IEEE.

Ng, S. T., Skitmore, M., & Wong, K. F. (2008). Using genetic algorithms and linear regression analysis for private housing demand forecast. Building and Environment, 43(6), 1171-1184.

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.

Núñez Tabales, J. M., Caridad y Ocerin, J. M., & Rey Carmona, F. J. (2013). Artificial neural networks for predicting real estate prices. Revista de Metodos Cuantitativos para la Economia y la Empresa, 15, 29-44.

Jiang, L., Phillips, P., & Yu, J. (2014). A new hedonic regression for real estate prices applied to the Singapore residential market.

Onder, Z. (2000). High inflation and returns on residential real estate: evidence from Turkey. Applied Economics, 32(7), 917-931.

Öztürk, N., & Fitöz, E. (2012). Türkiye'de konut piyasasının belirleyicileri: ampirik bir uygulama. Uluslararası Yönetim İktisat ve İşletme Dergisi, 5(10), 21-46.

Özkan, G., & Yalpir, S. (2005). Taşınmaza ekonomik bakış ve değerlendirmesi. TMMOB Harita ve Kadastro Mühendisleri Odası, 10.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ...& Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.

Peterson, S., & Flanagan, A. (2009). Neural network hedonic pricing models in mass real estate appraisal. Journal of Real Estate Research, 31(2), 147-164.

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: a review of valuation methods. Journal of Property Investment & Finance, 21(4), 383-401.

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.

Rossini, P. (1997). Artificial neural networks versus multiple regression in the valuation of residential property. Australian Land Economics Review, 3(1), 1-12.

Rumelhart D. E., McClelland J. L., 1986, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, (Vol.1), MIT Press,Cambridge, Massachusetts

Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. bioinformatics, 23(19), 2507-2517.

Saraç, E. (2012). Yapay sinir ağları metodu ile gayrimenkul değerleme (Doctoral dissertation, İstanbul Kültür Üniversitesi/Fen Bilimleri Enstitüsü/İnşaat Mühendisliği Anabilim Dalı).

Sarip, A. G., Hafez, M. B., & Daud, M. N. (2016). Application of fuzzy regression model for real estate price prediction. Malaysian Journal of Computer Science, 29(1), 15-27.

Sayer, J., & Moohan, J. (2007). An analysis and evaluation of hedonic price valuations in local leasehold office markets. In 13th Conference of the Pacific Rim Real Estate Society (pp. 21-24).

Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. Expert Systems with Applications, 36(2), 2843-2852.

Stepnowski, A., Moszyński, M., & Van Dung, T. (2003). Adaptive neuro-fuzzy and fuzzy decision tree classifiers as applied to seafloor characterization. Acoustical Physics, 49(2), 193-202.

Stock, J. H., & W Watson, M. (2003). Forecasting output and inflation: The role of asset prices. Journal of Economic Literature, 41(3), 788-829.

Stone, M. (1974). Cross- validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 111-133.

Tay, D. P., & Ho, D. K. (1992). Artificial intelligence and the mass appraisal of residential apartments. Journal of Property Valuation and Investment, 10(2), 525-540.

Teixeira, M. C. C., Caridad, J. M., & Ceular, N. (2010). Hedonic methodologies in the real estate valuation. In Mathematical Methods in Engineering International Symposium. Instituto Politécnico de Coimbra.

Tsatsaronis, K., & Zhu, H. (2004). What drives housing price dynamics: cross-country evidence.

Tsoukalas, L. H., & Uhrig, R. E. (1996). Fuzzy and neural approaches in engineering. John Wiley & Sons, Inc..

Ustundag, A., Cevikcan, E., & Kilinc, M. S. (2011). A Hybrid Fuzzy Risk Evaluation Model For Real Estate Investments. Journal of Multiple-Valued Logic & Soft Computing, 17(4).

Yayar, R., & Demir, D. (2014). Hedonic estimation of housing market prices in Turkey. Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, (43), 67-82.

Wang, X., Wen, J., Zhang, Y., Wang, Y. (2014). Real estate price forecasting based on SVM optimized by PSO. Optik-International Journal for Light and Electron Optics, 125(3), 1439-1443.

Weiss, S. M., & Kulikowski, C. A. (1991). Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems. Morgan Kaufmann Publishers Inc..

Wheaton, W. C. (1999). Real estate "cycles": some fundamentals. Real estate economics, 27(2), 209-230.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits (No. TR-1553-1). Stanford Univ Ca Stanford Electronics Labs.

Wilamowski, B. M. (2003, December). Neural network architectures and learning. In IEEE International Conference on Industrial Technology, 2003 (Vol. 1, pp. TU1-T12). IEEE.

Worzala, E., Lenk, M., & Silva, A. (1995). An exploration of neural networks and its application to real estate valuation. Journal of Real Estate Research, 10(2), 185-201

Zurada, J. M., Levitan, A. S., & Guan, J. (2006). Non-conventional approaches to property value assessment. Journal of Applied Business Research, 22(3), 1.

Invest in Turkey. (2016). Available at http://www.invest.gov.tr/enUS/sectors/Pages/RealEstate.aspx .

Global Property Guide. (2016). Available at http://www.globalpropertyguide.com/Europe/Turkey/Price-History-Archive/Strong-but-slower-house-price-growth-in-Turkey-127497

CURRICULUM VITAE

She was born in 1983 in Hatay. After completing her primary and high school education in Hatay, she entered Mersin University Computer Engineering Department in 1998. During her university years, she worked as a trainer in a private computer education institution and also gave private computer lessons. In the last year of the university, she managed the reporting and maintenance of the used “Micro” software in a store's IT department. After completing her university education, she continued her career in 2003 as a software specialist in a private company in Adana. She worked as a software specialist, project manager and analyst in private companies for 10 years. In addition, she developed software using many programming languages (e.g.: ASP.Net, c#, VB.net, PHP, HTML, JS, CSS, T-SQL ...) She also has developed software in different sectors (ERP, E-commerce, Health, CRM etc.). During this period, after receiving the training of Logo and Netsis software, she passed the exams and became a solution partner. Her softwares integrated into these systems are still in use. In addition, during this period, she worked as a contracted lecturer at Çukurova University Faculty of Education for one year. She has been working as a lecturer at Adana Alparslan Türkeş Science and Technology University since 2013. Besides, she also provides professional programming and database language training with software expertise. In 2019, she is planning to complete her master's degree in Industrial Engineering at Adana Alparslan Türkeş Science and Technology University.