Eskişehir Osmangazi Üniversitesi İİBF Dergisi
Ağustos 2018, C. 13, S. 2, 227 – 238.

Başvuru : 22.04.2018
Kabul : 18.07.2018

Examination of the Factors Affecting Household Rental Housing Demand Through Data Mining: The Case of Turkey

Kübra Önder[1]
Hüseyin Turgut[2]

Hanehalkının Konut Kira Talebine Etki Eden Faktörlerin Veri Madenciliği ile İncelenmesi

Öz

Konut, barınma açısından vazgeçilmez bir araç olmasına rağmen son yıllarda dünyada yaşanan ekonomik, kültürel ve teknolojik değişimlerin bir sonucu olarak konut, barınma temel ihtiyacını karşılama amacı dışında, gelecek için güvence kaynağı, yaşam tarzı ve zenginliğin en önemli göstergelerinden biri olmakla kalmamış toplum içindeki saygınlığın da ifadesi olmuştur. Bu durum konutun heterojen bir mal olmasına ve birçok özelliği içinde barındırmasına neden olmaktadır. Konutun sahip olduğu farklı özellikleri konutun piyasa değeri üzerinde ne derece etkin olduğu mikro iktisat teorisine dayalı hedonik fiyat modeli ile tahmin edilmektedir. Çalışmanın amacı, hanehalkının kiralık konut talebini etkileyen faktörleri Veri Madenciliği süreç ile analiz etmektir. 2015 yılı Hanehalkı Bütçe Anket verilerinin kullanıldığı çalışmada, 2341 hane verisi çalışmaya kaynaklık etmiş ve 49 farklı veri başlığı temel alınmıştır. Çalışma sonucunda, Veri Madenciliği yöntemlerinden Karar Ağacı Algoritmasının en iyi sonuç verdiği tespit edilmiştir.

Anahtar Kelimeler: Konut Piyasası, Hedonik Fiyatlama, Veri Madenciliği, Karar Ağaçları Yöntemi.

Examination of the Factors Affecting Household Rental Housing Demand Through Data Mining

Abstract

Houses are an irreplaceable tool for the need for shelter However, as a result of the global economic, cultural and technological changes encountered in recent years they have become a source of assurance for the future and one of the most significant indicators of life-style and wealth and social prestige as well as meeting the basic need for shelter. This causes house to become a heterogeneous good and to involve many characteristics. The effectiveness level of different characteristics of the house on the market value of the house is predicted using hedonic pricing model based on micro-economic theory. The aim of the study is to analyze the factors affecting rental housing demand of households through data mining methods. 2341 household data has been referenced and 49 data title was selected as the basis ofthe study in which Household Budget Survey data of 2015 has been used. As a result of the study, it has been determined that Decision Tree algorithm which is one of the Data Mining methods yielded the best result.

Keywords: Housing Market, Hedonic Pricing, Data Mining, Decision Trees Method.

1. Introduction

Housing meets the need for shelter, which is among the basic requirements of people. Sheltering is not just an individual's protecting himself/herself against the environmental conditions but also is a concept to which various meanings are attributed. This concept includes the meanings such as protecting the privacy of the family and the 'honor' of the women in the social value and life in which the individual lives (Ghannam, 1997). Starting from the definition, the significance of the position of housing in human life is at an indisputable level. Besides its existential importance, the economic, legal, and technical dimension with its numerical weight also increases its significance. Although the Constitution's "housing and sheltering rights" regulation is generally based on home ownership, of course, not all people have housing. Rapid population growth, the migration of people from rural to urban areas as a result of the mechanization of agriculture, and the increase in the number of refugees and immigrants

[1] Dr.Öğr.Üyesi, Mehmet Akif Ersoy Üniversitesi, İİBF, İktisat Bölümü. konder@mehmetakif.edu.tr, Yazar ORCID bilgisi: http://orcid.org/0000-0003-3537-3635.
[2] Öğr.Gör., Mehmet Akif Ersoy Üniversitesi Teknik Bilimler, MYO, Bilgisayar Teknolojileri Bölümü. hturgut@mehmetakif.edu.tr. Yazar ORCID bilgisi: http://orcid.org/0000-0001-8877-3179.

increase the need for housing day by day. As of 2016, although the number of houses sold for the first time across Turkey reaches 50.3% share in total housing sales, with approximately 72 thousands, one of the four people in Turkey are still tenants (TUIK, 2018a). Hence, individuals' housing demands are met by either buying or renting. Although individuals' demand for housing has an important weight in the economy, it also constitutes a significant place in household consumption expenditures. According to the the data of 2016 household budget survey published by TUIK, housing and rent expenditures had the highest share in total consumption expenditures with a rate of 25.2% in overall Turkey (TUIK, 2018b) When the sector having such an economical importance is examined in terms of the housing ownership status; it is seen that 59.7% of the individuals are homeowners in the residence they live, 24.4% are the tenants, 1.5% reside in the lodgings and 14.4% are in the other group (TUIK, 2018c). When the rented housings are taken into consideration; the rented residence may not only meet the need for housing but also be a property in which the individual acts with a demonstration and/or snob effect. Therefore, the price of rental housings is not only affected by the supply and demand relation. The environmental and the other characteristics of the housing with the characteristics of the individuals (social status, income etc.) also affect the price of rental housing. There are many factors affecting the price of the house, which is a heterogeneous commodity. There are several approaches to determine the factors which have an impact on the price of the housing and to create the best combination. These approaches have usually been combined into two main frameworks, traditional and advanced methods[3] (Pagourtzi et al. 2003). The hedonic price method is widely used in the literature (Goodman, 1978; Harrison et al., 1978; Steppard, 1999).

The hedonic price model, based on the consumer theory of classical economics, states that each of the individual qualities, and characteristics of each commodity provides consumers with different satisfaction or utility level, taking the assumption that goods are heterogeneous as a basis. In a hedonic approach based on the assumptions of full competition and full information, the characteristics possessed by a good fulfill the different needs of the consumer, and as a result of the consumption of each characteristic, the utility level it provides for the consumer also changes. Hedonic pricing is simply a kind of valuation system, and it can be said that the value of a good is measured over the value of the characteristics it has. In other words, it is the way to estimate the implicit price relation between prices, and the features of goods sold in the differentiated product market (Hidano, 2002; Shimizu et al., 2010: 354). Although Haas (1922) was the first person who used the method, the most significant contribution to the theory was made by Lancaster (1966) and Rosen (1974). The hedonic price method explores the marginal effect of each variable that is effective in determining the sale or rental value of a real estate and the decisiveness and contribution of those variables to the rental or sale price of the real estate. Although the hedonic price model has advantages such as being a more transparent when evaluated in the framework of classical index theory (Koev, 2003), this approach has been criticized for the possible problems. These problems arise from the estimations such as basic model assumptions and supply and demand definition, market imbalance, the choice of independent variables and functional form of the hedonic equation,

[3] Traditional methods are comparable methods, investment/income methods, multiple regression analysis method and stepwise regression method. Advanced methods, on the other hand, are hedonic price method, artificial neural networks (ANN), spatial analysis methods, fuzzy logic and autoregressive integrated moving average method. For detailed information, please see Pagourtzi et al., 2003.

and market segmentation (Fan et al. 2006; Sivitanides 1997; Vries et al. 2009). Therefore, since the 1990s, the artificial neural networks (ANN) method, which is a flexible regression approach in predicting housing prices, has started to be used quite frequently. When the literature is examined; it's seen that there are studies addressing the hedonic price and ANN approach together, and in all of these studies it has been emphasized that ANN method is better than hedonic price method in estimation performance (Nghiep and Al 2001; Limsombunchai et al., 2004, Selim, 2009; Peterson and Flanagan, 2009, Morano and Tajani, 2013; Önder, 2017). In real estate valuation, besides the ANN and hedonic price method, data mining methods have also been used and the obtained results have been compared (Jaen, 2002; Fan et al., 2006; Del Cacho, 2010; Ganet al., 2015). Data Mining is the process of parsing and scanning the useful information and processing huge volume of data emerged by the information age. It is almost impossible to analyze and transform this huge volume of data to useful information by using statistical methods. Knowledge discovery or data mining is becoming a support in this point. It is enabled to with this method in current competition environment. Data mining method is less complex and easy to realize in comparison to other methods. Besides, this method differs from other methods as it proposes unprecedented results.

In these studies, it has been emphasized that the best result of these methods belongs to the decision tree algorithm. Unlike other studies in national and international literature, data mining processes and methods were used instead of hedonic price model and ANN approach in this study. The aim of the study is to determine the existence of the correlation among the answers given to the questions that the Turkish Statistical Institute (TUIK) has used for the 2015 Household Budget Survey in terms of real estate rental value, using the data mining method. Furthermore, through the determination of the existence of the correlation among the data, it is also helpful in eliminating the lack of data that may come up in the future. As a result, the study consists of three parts. In the first part, the dataset and method used in the study are described. In the second part, the findings of the research are presented, and in the last part, evaluations were made about the subject in the light of the findings obtained.

2. Data Set and Method

2.1. Data Set

The scope of this study, in which 2015 Household Budget survey data are used, consists only of the dataset which belongs to the household residing as a tenant. Therefore, date with 2341 household are the sources of this study. 49 different data headings were used as a base in the analysis of the correlation between the rental value and the physical, environmental and structural characteristics of the housing and the socio-economic characteristics of the household. The data headings and descriptions included in the study are presented in Table 1.

Table 1: Data Used in the Study

VARIABLES	VALUE
Rent amount	Less (0-450 TL) Medium (451-900 TL) High (901- ... TL)
m2 of housing	Value
Duration of stay in the housing	Value (month)
Housing type	Apartment / condo
Rental type	Furnished / unfurnished
The employment tool	Yes / No

Generator	Yes / No
Cablecast	Yes / No
Underground garage	Yes / No
Microwave oven	Yes / No
Motorcycle	Yes / No
Automatic washing machine	Yes / No
Elevator	Yes / No
Garden	Yes / No
Balcony	Yes / No
Dishwasher	Yes / No
A refrigerator	Yes / No
Washing Machine	Yes / No
Children's playground	Yes / No
Deep freeze	Yes / No
Continuous hot water	Yes / No
Natural gas ownership	Yes / No
Security officer system	Yes / No
Carpet washing machine	Yes / No
Second housing ownership	Yes / No
Internet subscription	Yes / No
Game console	Yes / No
Panel TV	Yes / No
Telephone line	Yes / No
Satellite antenna	Yes / No
Video camera	Yes / No
swimming pool	Yes / No
At least the number of parents in HanDa	(see Annex-1 "a")
At least the number of children	(see Annex-1 "a")
Household size	(see Annex-1 "a")
Primary fuel type for heating purposes	(see Annex-1 "b")
Fuel for hot water	(see Annex-1 "b")
Fuel used for cooking	(see Annex-1 "b")
Basic exothermic shape	(see Annex-1 "c")
Number of air conditioners	Quantity (units)
Number of computers in the house	Quantity (units)
Shopping access difficulty	Yes / No
Difficult access to banking services	Yes / No
Difficulty accessing postal services	Yes / No
Difficulty in accessing health services	Yes / No
Difficulty in accessing compulsory education service	Yes / No
Number of cars	Quantity (units)
Household expenditure	Value (TL / month)
The age of the building	Value
Monthly income	Low (0-1400) Medium (1401-3000) High (3001- ..)

Considering minimum wage of the year 2017, household monthly income used in the study was classified as low, middle and high whereas the variable for the age of building was classified as new (0-10 years), middle (11-25 years), old (26-45 years) and very old (46+ years). In this classification, the depreciation index which is published by the Revenue Administration (GIB, 2017) is used. Since household size and the relatives status of the individuals living in the housing will determine the numbers of the rooms in the housing to be rented and hence the rental value of dwelling, these variables were transformed by the authors and included in the study (Annex-1 "a"). The primary fuel type for heating purposes and the type of fuel used in getting hot water and cooking were grouped as natural gas, electricity, fuel oil, solar, coal, LPG, wood, thermal, turf, none and other (Annex-1 "b"). And the basic heating type of the house was defined as combi boiler, air conditioner, central heating, stove, none and other (Annex-1 "c"). In the survey, the time spent in the housing was taken as annual or monthly. In the study, annual data were converted into monthly data to ensure uniformity. Monthly rental fee variable was classified as low, medium and high considering the frequency values.

2.2. Method

Data mining method was used in the study with the aim of defining the correlations among the data obtained from the TUIK 2015 household budget survey.

Accumulated and accumulating data in large databases are forming significant or insignificant data stacks day by day. Obtaining meaningful information from such a large stack of information is an additional benefit. Data mining is the only way to extract valuable information from such complex and massive data (Lausch et al., 2015). The extraction period of these complex and massive data takes place in five stages, which are shown below (Shearer, 2000):

1. Identifying the problem,
2. Preparing the data,
3. Establishing and evaluating the model,
4. Using the model,
5. Monitoring the model.

The most important condition of success in data mining studies is the problem identification stage. At this stage, the purpose of the study should be clearly expressed. Another important step in data mining is the data preparation stage. This stage consists of the sub-steps of collecting, integrating, cleaning and transforming the data. Processes such as determining the sources of collected data, eliminating the deficiencies or errors existing in the dataset, and signifying the insignificant data are carried out at this stage. The stage of establishing the model, on the other hand, is the process of determining the most successful model that is appropriate to solve the problem using prepared data. All models can be tried or the data preparation stage can be repeated to find the most successful model. The successful model obtained by using the model can be used either directly or as part of a large workflow. In the stage of monitoring the model, the last stage, the entire system and the model are monitored. At this stage, the system can be updated with new data if needed (DuMouchel, 1999).

Data mining, which enables us to reach valuable information from complex and massive data, can also be defined as an analytical tool and computer software package categorizing data and helping to determine the correlations among the data. Specific data mining software includes many models such as clustering, linear regression, neural networks, Bayesian networks, visualization and tree-based models. In today's data mining technology, contrary to old methods, high-speed computers are used to detect trends and correlations in large datasets in a short period of time and reveal unknown trends (Turgut, 2012). This method can be grouped into three main headings. These are classification, clustering and association rules and sequential patterns. In addition to methods such as classification and regression, clustering, artificial neural networks, decision trees algorithms are also defined as application models of these methods (DuMouchel, 1999).

Decision trees algorithm is easier to mount and more understandable compared to other methodologies (Chien et al., 2008). Due to these advantages of the algorithm, it is the most frequently used method in data mining approaches. In decision tree algorithms, like other classification algorithms used in data mining, the data can be grouped into two sub-groups as training and test data. The first stage in decision trees implementation is the learning step. While training data is used to form the model, the test data are used for analysis and verification of the correctness of the generated model.

3. Implementation Process and Research Findings

In the selection of data mining software, WEKA software was used with performance/cost evaluation criterion (Witten et al., 2000). All data was converted to csv or arff format in compliance with WEKA software. In the implementation process, the stages of identifying the problem, preparing the data, establishing and evaluating the model, using the model and monitoring the model were applied, respectively.

3.1. Identifying the Problem

The problem of determining what kind of relationship there is between the data of rental expenditures and the data on physical and environmental characteristics of housing and social and economic values of household, and if there is any, how this relationship can be modeled successfully was identified. To solve this problem, B micro-data obtained from TUIK was used.

3.2. Preparing the Data

The data obtained from the TUIK was examined and more than 100 data headings were determined. Of all data, particularly the ones belonged to households who were the owners of the housing and those who didn't pay rent to reside in the housing were removed from the dataset. In the data preparation stage, in compliance with the WEKA used in data mining practice, headings were converted to a style without spaces and without having any Turkish characters. The housing type values were grouped as a house and a flat, and 98 data not belonging to this group were extracted from the exception sample dataset. As a way of renting, the data were subjected to transformation as furnished and unfurnished. From the year and month values given as the dwelling time in the housing, the values were converted into the numeric format for the definition of the software, and each value in each instance was standardized as the month. The construction date of the building was grouped as shown in Table 1 and converted into nominal values. The exceptional values found in the heating type have been deleted. The data in the form of transportation difficulty were reduced to 3 different groups. Exceptional examples such as sauna and jacuzzi have been deleted. Many values were subtracted from the dataset in the exceptional case and the data giving the quantity in the data definition were converted to the nominal values according to the grouping of either yes or no. The dataset was defined as 2385 samples and 66 headings after all these processes. 49 different headings and 2341 samples, affecting the rental expenditure value converted by the "selection attribute" process found in WEKA software were prepared as a usable dataset.

3.3. Establishing and Evaluating the Model

In the course of establishing the model, model establishment trials were carried out with decision tree algorithms out of classification algorithms by the WEKA process over the dataset. The cross-validation value was set at 10 during the applications. In the study, the dataset was divided into 10 equal groups and a correlation between 9 datasets and other group was established. The values obtained through the application are presented in Table 2.

Table 2: Successful Result Values of Successful Algorithms

ALGARORITHM NAME	ACCURATELY ESTIMATED "HIGH"	ACCURATELY ESTIMATED "MEDIUM "	CORRECTLY ESTIMATED "LOW"	THE NUMBER OF ACCURATE ESTIMATE	ACCURACY RATE
Decisionstump	0	678	918	1596	68,176
Hoeffdingtree	77	514	1142	1733	74,0282
J48	45	497	1157	1699	72,5758
Lmt	52	545	1204	1801	76,9329
Randomforest	49	542	1209	1800	76,8902
Randomtree	48	460	1081	1589	67,877
Reptree	37	524	1128	1689	72,1487

When Table 2 is examined, it was determined that the Logistic Model Tree (LMT) algorithm with the highest accurate estimate and accuracy is suitable and applicable. The accuracy and false estimate rates of the LMT algorithm are presented in Table 3.

Table 3: The Summary of the Accuracy and False Estimate Rates of the LMT Algorithm

	TP	FP	PRECISION	RECALL	F-MEASURE	MCC	ROC	PRC	CLASS
	0,388	0,012	0,667	0,388	0,491	0,487	0,939	0,562	high
	0,687	0,183	0,657	0,687	0,672	0,499	0,841	0,678	medium
	0,851	0,248	0,84	0,851	0,846	0,606	0,893	0,929	low
Weighted Avg.	0,769	0,213	0,768	0,769	0,766	0,563	0,878	0,823	

When Table 3 is examined, the positive accurate estimate values for each variable are 38.8% for "high", 68.7% for "medium" and 85.1% for "low". Positive false estimate values for each variable are 1.2% for "high", 18.3% for "medium" and 24.8% for "low". The correct estimate value for all variables is 76.9%. When these values are adapted to 134 "high", 793 "medium" and 1414 "low" values in the dataset, they have the distribution indicated in Table 4.

Table 4: Estimation Distribution of LMT Algorithm

Classification ==>	**a**	**B**	**c**	Total
a- high	**52**	77	5	134
b- medium	23	**545**	225	793
c- low	3	207	**1204**	1414

When Table 4 is investigated, the estimation distribution of the classification is seen. 1801 (52+454+1204) out of 2341 (134+793+1414) samples were accurately estimated.

2.4. Using the Model

The decision tree model, obtained by the LMT algorithm, for the correlation between the rental situation and other headings is shown in Table 5.

Table 5: Obtained LMT Algorithm Model

Logistic model tree ------------------ : LM_1:61/61 [2341] Number of Leaves : 1 Size of theTree : 1 LM_1:
Class high : 0.12 + [parentsinhousehold]* -0.48 + [householdsize]* -0.24 + [building_age=old]* 0.66 + [building_age=very old]* 1.1 + [primary heating= air conditioner]* -0.29 + [primary_heating=stove]* -2.79 + [fuel_1=natural gas]* 1.33 + [fuel_1=coal]* -1.8 + [fuel_1=thermal]* 1.8 + [fuel_hotwater=electricity]* 0.2 + [fuel_hotwater=solar]* -2.69 + [cable=yes]* 0.37 + [elevator=yes]* 0.29 + [swimming pool=no]* -2.67 + [generator=yes]* 0.82 + [telephone line=yes]* 0.31 + [number of telephone line= many]* -2.46 + [number of computer]* -0.13 + [internet subscription=no]* -0.55 + [panel TV=no]* -0.62 + [video_camera=yes]* 0.26 + [satellite antenna=no]* 0.26 + [freezer=yes]* 0.22 + [dishwasher=no]* -0.17 + [automatic_washing_machine=no]* -0.8 + [carpet_washing_machine=yes]* -0.64 + [the number of air conditioning]* 0.54 + [game_console=yes]* 0.41 + [the number of car]* -0.58 + [montly_income_level=high]* 0.52 + [monthly_income_level=low]* -0.86
Class middle : -1.38 + [housing_type=apartment]*1.46+ [Renting_type=furnished] *0.29+ [primary_heating=combi]*-0.07 +[primary_heating=central]*-.04+ [primary heating=other]*-0.17+ [primary_heating=stove]*0.49+ [primary_heating=no]*-1.27+ [fuel_1=coal]* -0.1+ [fuel_1= electricity]* 0.26 + [fuel_1= wood]* -0.71 + [fuel= LPG] * 0.69 + [fuel_1= solar]* 0.99 + [fuel_cooking= natural gas]* 0.13 + [fuel_cooking=electricity]* -0.87 + [fuel_cooking=fuel-oil]* -2.31 + [fuel_hotwater= natural gas]* -0.08 + [fuel_hotwater = LPG]* 0.16 + [continues_hotwater =no]* -0.25 + [shopping_difficulty=difficult]* -1.84 + [postal_service_difficulty= difficult]* 0.26 + [healt_service_difficulty= easy]* 0.15 + [healt_service_difficulty= difficult]* -0.96 + [compulsory_education_difficulty= medium]* -0.06 + [compulsory_education_difficulty=difficult]* 0.93 + [security_personnel_system=no]* 0.19 + [balcony=yes]* 0.38 + [garden= no] * 0.24 + [playground= no] * -0.35 + [internet_subscription= no] *- 0.1+ [microwave oven= no] * 0.19 + [dryer= yes]*0.21+ [motorcyle= yes]*-0.25+ [monthly_income_level= middle]*0.09+ [monthly_income_level= high]*-0.08+ [monthly_income_level=low]*-0.2
Class low : -0.09 + [min.child_inhousehold]* 0.03 + [parents_inhousehold]* 0.3 + [building_age= old]* -0.09 + [building_age= very old]* -0.16 + [housing square_meter]* -0.01 + [primary_heating= stove]* 1.35 + [primary_heating= no]* 1.28 + [fuel_1= coal] * 0.79 + [fuel_1= thermal] * -0.64 + [fuel_1= wood] * 0.27 + [fuel_cooking= electricity]* 0.59 +[fuel_hotwater=natural gas]* 0.16 + [gas_ownership= no]* 0.15 + [continous_hotwater=no]* 0.26 + [shopping_difficulty= difficult]* 2.92 + [banking_difficulty=easy]* -0.57 + [health_service_difficulty= difficult]* 1.85 + [second_housing_ownership=yes]* 0.24 + [cable TV=yes]* -0.78 + [elevator=yes]* -0.41 + [closed_garage= yes]* -0.48 + [swimming_pool= no]* 1.75 + [security_personnel_system= no]* -0.21 + [the_number_o_fcomputers]* -0.21 + [panel_tv=no]* 0.28 + [satellite_antenna= no]* -0.18 + [refrigerator= no]* 0.56 + [dishwasher= no]* 0.26 + [carpet_washing_machine= yes]* 0.15 + [game_console= yes]* -0.67 + [number_of_cars]* 0.52 + [service vehicle= yes]* -0.28 + [motorcycle= yes]* 0.5 + [monthly_income_level= high]* -0.83

The model obtained by the LMT algorithm, which was successful in the data mining process applied in the solution of the problem, has created three different equations. These equations created according to this model belong to high, medium and low classes. Each equation consists of numerical values. The largest of the numerical results obtained is the estimated value of that sample. The rental values of the equations are classified into three sub-groups. The variables in three equations are used quantitatively or qualitatively. If variable in the equation is quantitative, it is added to the sum by multiplying the numerical value with the related coefficient. If variable in the equation is qualitative, in this case, it is added to the sum by multiplying the related coefficient with 1 if the qualitative proposition is true and it is added to the sum by multiplying the related coefficient with 0 if the qualitative proposition is false.

Equations for each class are calculated separately. As a result of three separate equations, the rental value is high if the class with the highest value is high; the rental value is middle if it is middle; the rental value is low if it is low.

2.5. Monitoring the Model

The applicability of this generated model was tested and no probability which might create instability in the datasets was observed. Thus it was not necessary to repeat the work.

Using the dataset created with the data obtained by TUIK's field survey, it has been proved that there is a correlation between the rental value and other values of the housing. It has been determined that this correlation was obtained with the LMT algorithm with an accuracy of 76.9329%, and the most suitable model in estimating rent value of a rented housing is LMT algorithm.

3. Conclusion

The fact that the housing is a heterogeneous commodity and accommodates many properties has an effect on the rental values of the housing. Since housing is priced as a whole, these properties cannot be marketed clearly. This does not mean that the characteristics of the housing do not have an effect on its rental value. With the hedonic price model, it is possible to calculate the effect that the different characteristics of the products have on the price of the product. There is a non-linear correlation between the housing characteristics, which have a significant effect on its rental value, and the rental value of the housing. The use of linear approaches in estimating the rental value of housing is deterring confidence in the estimation results (Fan et al., 2006; Malpezzi, 2003). For this reason, hedonic price, ANN and data mining approaches have been used in the literature. In this study, the existence of the correlation between the rental value and the socio-economic structure of the household, the structural and environmental characteristics of the housing were analyzed, using the data mining method newly introduced in national literature; In the study where 2341 household data were used, 49 different data headings were taken as a basis. It has been determined that there is a correlation between the data headings in the dataset generated by TUIK data. The existence of these correlations among the data indicates that the data mining process can also be applied to these data. As with all other data mining applications, this process can be applied in the estimation of missing or new data. It is envisaged that the accuracy value will increase more by applying data collection to a specific region.

This study has shown that which characteristics the consumers who aim for utility maximization consider when demanding the housing and that the rental value can be determined by the characteristics of the housing and the characteristics of the tenant household. Moreover, it also identified the methods used by the tenant, the homeowner and the intermediaries to determine the rental price of the housing. This study, which has been successful in estimating or determining the rental value, will open up new web applications in the future, increasing its applicability, especially in the virtual world. That the construction and real-estate firms build housings considering the housing preferences of consumers will enable efficient use of scarce resources. Since consumer preferences may vary by residential locations and time, that such studies are conducted at different times and in different locations may provide guidance for construction and real estate companies operating in the region concerned. It is thought that the study will also contribute to the future micro and macro level studies on the housing market.

References

Brian, T. Ratchford (1975), "The New Economic Theory of Consumer Behavior: An Interpretive Essay", *Journal of Consumer Research*, Vol. 2, No. 2: 65-75, https://doi.org/10.1086/208617

Chien, Chen-Fu; Chen, Li-Fei (2008), "Data Mining to Improve Personnel Selection and Enhance Human Capital: A Case Study in High-TechnologyIndustry", *Expert Systems with Applications*, Vol. 34: No. 1: 280-290.

Colin, Sheare (2000), "The CRISP-DM Model: The New Blueprint for Data Mining", 13.

Del Cacho, Carlos (2010), "A Comparison of Data Mining Methods for Mass Real Estate Appraisal", MPRA, *https://mpra.ub.uni-muenchen.de/27378/1/A_comparison_of_data_mining_methods_for_mass_real_estate_appraisal.pdf*, (Erişim: 02.11.2017).

DuMouchel, William (1999), "Bayesian Data Mining in Large Frequency Tables, with an Application the FDA Spontaneous Reporting System", *The American Statistician*, Vol. 53, No. 3: 177-190.

Fan, Gang-Zhi; Ong, Seow Eng; Koh, Hian Chye (2006), "Determinants of House Price: A Decision Tree Approach", *Urban Studies*, Vol. 43, No. 12: 2301-2315.

Gan, Victor; Agarwal, Vaishali; Kim, Ben (2015), "Data Mining Analysis and Predictions of Real Estate Prices", *Issues in Information Systems*, Vol. 16, Issue IV: 30-36

Goodman, Allen C. (1978), "Hedonic Prices, Price Indices and Housing Markets", *Journal of Urban Economics*, Vol. 5, No. 4: 471-484.

Grannam, Farha (1997), Re-imagining the Global: Relocation and Local Identities in Cairo, içinde A.Öncü ve P. Weyland (der.), Space, Culture and Power: New Identities in Globalizing Cities (ss.119-139). Londra ve New Jersey: Zed.

Haas, G. C. (1922), A statistical Analysis of Farm Sales in Blue Earth County, Minnesota, As a Basis For Farm Land Appraisal, Yayımlanmamış Yüksek Lisans Tezi, Minnesota Üniversitesi, ABD.

Harrison, David; L. Rubinfeld, Daniel (1978), "Hedonic Housing Prices and the Demand for Clean Air", *Journal of Environmental Economics and Management*, Vol. 5, No. 1: 81-102.

Hidano, Noboru (2002), *The Economic Valuation of Environment and Public Policy: A Hedonic Approach*, Edward Elgar, Massachosetts.

Jaen, Ruben D. (2002), "Data Mining: An Empirical Application in Real Estate Valuation", In FLAIRS Conference http://www.aaai.org/Papers/FLAIRS/2002/FLAIRS02-062.pdf , (Erişim: 28.01.2018).

Koev, Eugen (2003), Combining Classification and Hedonic Quality Adjustment in Constructing a House Price Index, Licentiate Thesis Department of Economics University of Helsinki, Finland.

Lancaster, Kelvin J. (1966), "A New Approach to Consumer Theory", *Journal of Political Economy*, Vol. 74, No. 2: 132–157.

Lausch, Angela; Schmidt, Andreas; Tischendorf, Lutz (2015), "Data Mining and Linked Open Data–New Perspectives For Data Analysis in Environmental Research", *Ecological Modelling,* Vol. 295: 5-17.

Malpezzi, Stephen (2003), Hedonic Pricing Models: A Selective and Applied Review, in: T.O'Sullivan and K. Gibb (Eds) Housing Economics and Public Policy, 67–89, Malden, MA: Blackwell Science, https://doi.org/10.1002/9780470690680.ch5

Nguyen, Nghiep; Cripps, Al (2001), "Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks", *Journal of Real Estate Research,* Vol. 22, No. 3: 313-336.

Önder, Kübra (2017), *Current Approaches in Education and Economics*, (Ed.)Hasan GÖKSU, *"Comparative Analysis of The FactorsThat Affect House Prices Using Hedonic Pricing Model And Artificial Neural Network (ann) Method: A Case Study in Burdur Province",* Bölüm VI: 75-94, SRA Academic Publishing, United States of America.

Pagourtzi, Elli; Assimakopoulos, Vassilis; Hatzichristos, Thomas; French, Nick (2003), Real Estate Appraisal: A review of Valuation Methods", *Journal of Property Investment & Finance*, Vol. 21, No. 4: 383-401.

Pierluigi, Morano; Francesco, Tajani (2013), "Bare ownership evaluation. Hedonic price model vs. artificial neural network, International Journal of Business Intelligence and Data Mining, Vol. 8, Issue 4: 340-362. https://doi.org/10.1504/IJBIDM.2013.059263

Ratcford, Brian T. (1975), "The New Economic Theory of Consumer Behavior: An Interpretive Essay", *Journal of Consumer Research*, Vol. 2, issue 2: 65-75.

Revenue Administration (GIB) (2017), Aşınma Paylarına İlişkin Oranlar Cetveli, https:// intvd.gib.gov.tr/ 2014_Emlak_Arsa/pdf/AsinmaPaylarinaIliskinCetvel.pdf, (Erişim: 02.12.2017)

Rosen, Sherwin (1974), "Hedonic Prices and Implicit Markets: Product Differentiation In Pure Competition", *Journal of Political Economy*, Vol. 82, No. 1: 35-55, http://agecon2.tamu.edu/people/faculty/cappsoral/agec%20635/Readings/Hedonic%20Prices%20and20Implicit %20Markets%20Product%20Differentiation%20in%20Pure%20Competition.pdf, (Erişim: 02.01.2018).

Selim, Hasan (2009), "Determinants of Houseprices in Turkey: Hedonic Regression Versus Artificial Neural Network", *Expert Systems with Applications*, Vol. 36, Issue 2, Part 2: 2843-52, https://doi.org/10.1016/j.eswa.2008.01.044

Shearer, Colin (2000), "The CRISP-DM Model: The New Blueprint for Data Mining", *Journal of Data Warehousing*, Vol. 5, Konu, 4: 13-22.

Sheppard, Stephen (1999), "Hedonic Analysis of Housing Markets", *Handbook of Regional and Urban Economics*, Vol. 3: 1595-1635.

Shimizu, Chihiro; Takatsuji, Hideoki; Ono, Hiroya; Nishimura, Kiyohiko G. (2010), "Structural and Temporal Changes in the Housing Market and Hedonic Housing Price Indices: A Case of the Previously Owned Condominium Market in the Tokyo Metropolitan 94 Area", *International Journal of Housing Markets and Analysis*, Vol. 3, Issue, 4: 351-368, doi: 10.1108/17538271011080655

Sivitanides, Petros (1997), "The Rent Adjustment Process and the Structural Vacancy Rate in the Commercial Real Estate Market", *Journal of Real Estate Research,* Vol. 13, No. 2: 195-209.

Steven, Peterson; Albert, Flanagan (*2009*), "Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal", *Journal of Real Estate Research*, Vol. 31, No. 2: 147-164.

Turgut, Hüseyin (2012), Veri Madenciliği Süreci Kullanılarak Alzheimer Hastalığı Teşhisine Yönelik Bir Uygulama, Yüksek Lisans Tezi, Isparta: SDÜ Fen Bilimleri Enstitüsü.

TÜİK (2018a), Konut Satış İstatistikleri, http://www.tuik.gov.tr/PreTablo.do?alt_id=1056, (Erişim: 14.01.2018).

TÜİK (2018b), Tüketim Harcaması Türleri, http://www.tuik.gov.tr/VeriBilgi.do?alt_id=1012, (Erişim: 14.01.2018).

TÜİK, (2018c), Hanehalkı Bütçe Anketi Tüketim Harcaması Sonuçları, http://www.tuik.gov.tr/VeriBilgi.do?alt_id=1012, (Erişim: 14.01.2018).

Visit, Limsombunchai; Christopher, Gan; Minsoo, Lee (2004), "House Price Prediction: Hedonic Price Model vs. Artificial Neural Network", *American Journal of Applied Sciences*, Vol. 1, No. 3: 193-201.

Vries, Paul de; de Haan, Jan; der Wal, Erna van; Mari´en, Gust (2009), "A House Price Index Based on the SPAR Method", *Journal of Housing Economics*, Vol. 18, Konu, 3: 214-223, https://doi.org/10.1016/j.jhe.2009.07.002

Wheaton, C. William, (1974), "A Bid Rent Approach to Housing Demand", *Journal of Urban Economics*, Vol. 4, Issue 2: 200-217, https://dspace.mit.edu/bitstream/handle/1721.1/63384/bidrentapproacht00whea.pdf?sequence=1 (Erişim: 14.01.2018).

Witten, Ian H; Frank, Eibe (2011), *Data Mining: Practical Machine Learning Tools and Techniques* (3rd edition), San Francisco: Morgan Kaufmann, United States of America.

Appendix-1 Explanation on Data Conversion

CONVERSION	EXPLANATION
a	Single-child nuclear family (at least 1 mother 1 father 1 child)
	Two-child nuclear family (at least 1 mother 1 father 2 children)
	Two or more-child nuclear family (at least 1 mother 1 father 3 children)
	Childless couple (at least 1 mother 1 father 0 child)
	Patriarchal or extended family (at least 1 mother 1 father 1 child)
	Single-adult family (at least 1 mother/father at least 0 child)
	People living together (at least 0 mother/father 0 child)
b	Natural gas, electricity, fuel-oil, solar, coal, LPG, wood, thermal, turf, none, other
c	Combi boiler, air-conditioner, central, stove, none, other