

USING SVR AND MRA METHODS FOR REAL ESTATE VALUATION IN THE SMART CITIES

A. U. Akar¹, S. Yalpir^{1*}

¹Department of Geomatics Engineering, Konya Technical University, 42100, Konya, Turkey
auakar@ktun.edu.tr, syalpir@ktun.edu.tr

KEY WORDS: Real Estate Valuation, Features, SVR, MRA, Valuation Model for Smart Cities and Urban.

ABSTRACT:

Determination of real estate value plays a very critical role in economic development and basic needs of people. Increasing demand for real estate together with population growth is making it difficult to determine real estate value. In applications where real estate is the main subject, such as urban activities, smart cities and urbanization, urban information system and valuation systems, model-based value estimations are essential for effective land/real estate policy. The type of real estate and impact degree of features depending on the type should be known as well as value estimation. It will be beneficial to follow a method that both determines the real estate value and factor impact degree. With the studies to be carried out using such methods, both region-specific valuation models can be created and the model is established with the optimum variable. This paper aimed to determine real estate value by using Support Vector Regression (SVR) and Multi Regression Analysis (MRA) methods for effective real estate management. Besides, both methods were examined by revealing the impact degrees of features that affect the value. The methods were applied to 319 parcels in Konya. For each parcel, 31 land features and market values were collected. The parcel data collected since 2018 were included in the models. From the results, the RBF-SVR model reached the highest R^2 value with 0.88, while the MRA model reached 0.86.

1. INTRODUCTION

Real estate or land valuation is the process of determining the value of real estate by considering its many features. Today, the rapid increase of population brings the need for land and subsequent building construction. Increasing demand for real estate as a result of population growth is also making it difficult to determine real estate value. In addition, this situation requires keeping a large number of real estate and its value in a system. Thus, accurate determining land types, features and values for valuation systems has great importance in terms of smart city applications, planned urbanization, detection of development regions and investment areas, taxation, and sustainable tax systems (Renigier-Biłozor et al., 2017). In this context, for value estimation, first, type of real estate and features depending on the type should be decided. In daily life, it is not possible to determine the exact value of any real estate. Because each real estate has many different features as of its location and use, and these features can vary from person to person in terms of quality and quantity (Eren et al., 1999). Moreover, used features and which of these features are more important are not known. As a solution, by using a scientific method for real estate valuation, the value of real estate can be determined more accurately and the effects of features on the value can be investigated.

The valuation process has been one of the important factors in the economies of the country due to its application in many areas such as taxation in municipalities, mortgage transactions in banks, expropriation, buying and selling. In addition, in smart cities, which have recently become very popular, play a major role in ensuring the sustainability of land management (Peris-Ortiz et al., 2017). While determining the value for real estate in these application areas, which are based on legislation, laws and regulations that require real estate valuation, value is estimated with different methods and real estate features. For this reason, the value of real estate varies as a result of different approaches. If it is desired to create a value estimation model

for all real estate in the region, several issues should be taken into account. There should be no differences in values according to the situation desired by the value estimator. All values and features in the region should be determined by mathematical methods. Therefore, valuation approaches that concern all real estate should be included in daily life. Recently, with the effect of rapid development in computer technology, new searches are still continuing. Artificial intelligence-based methods have been suggested as an alternative to MRA, which is frequently encountered in real estate value estimation in the literature, especially (Guan and Levitan, 1997; Guan et al., 2008; Król et al., 2007; McGreal et al., 1998; Peterson and Flanagan, 2009; Taffese, 2007; Worzala et al., 1995). Some of these methods are as follows: which is artificial intelligence methods; Artificial Neural Networks (ANN) (Peter et al., 2020), Fuzzy Logic (Yalpir and Ozkan, 2018), Rough Set Theory (Del Giudice et al., 2017) and Fuzzy AHP (Chang, 1996). In addition, which is machine learning optimization methods; Support Vector Machine (SVM) (Chen et al., 2017), Support Vector Regression (SVR) (Lee and Chen, 2016) and different regression models can be ordered as. On the other hand, to establish a real estate valuation model, it is necessary to find features that affect the value according to real estate type. For this, there are many studies discussed in the literature. In these studies, optimum criteria for the valuation model were determined by applying a survey to the actors (such as experts and citizens) (Unel and Yalpir, 2019). As another approach, the ordering of the most influential features or their effect degree was tried to be determined through the models used in value estimation (Chen et al., 2017; Baldominos et al., 2018; Doumpos et al., 2020).

This study's objective is to provide accurate, fairly and successfully determine real estate value by using SVR and MRA methods in mass valuation, land management and smart city applications. Besides, it is to determine the most effective features, by revealing the impact degree of features that affect the real estate value with scientific methods. The current study,

its target to provide solutions regarding the issues mentioned below.

- Deciding the way to be followed to determine on the features affecting the value.
- Providing support for decision-makers in applications (smart cities, urban information systems, real estate agencies, etc.) where real estate value is needed.
- To determine the impact degree of the features to be used in mass valuation.

In this context, for the central neighborhoods of Konya, the market value and features of parcels (lands) that are purchase-sale were collected. The parcel data collected since 2018 were included in the models. Real estate value estimation and effect degrees of features that affect the value were compared by utilizing SVR and MRA methods. Study results were examined statistically by comparing the model performance criteria. This paper mentions the features affecting the value of real estate, SVR and MRA in Section 2. The findings of the study are presented, interpreted and discussed in Section 3. Finally, the conclusions are expressed in Section 4.

2. MATERIAL AND METHODS

2.1. Features affecting the value of real estate

Accurate estimation of value is a crucial factor in real estate, which is the biggest investment tool for people. However, some topics need to be addressed to estimate its value correctly. There are numerous, complex and variable features that affect its value. Therefore, features used in the valuation should be examined by considering the study region and the optimum feature number for valuation model should be determined. In this study, the central neighborhoods of Selcuklu, Meram and Karatay, which are among the three central districts of Konya, were selected as study regions (Figure 1). Data collected from the study region have involved parcels (land qualified) that are subject to purchase and sale in market conditions. The spatial distribution of 319 parcels (red point: market samples) in Central Neighborhoods is given in Figure 1.

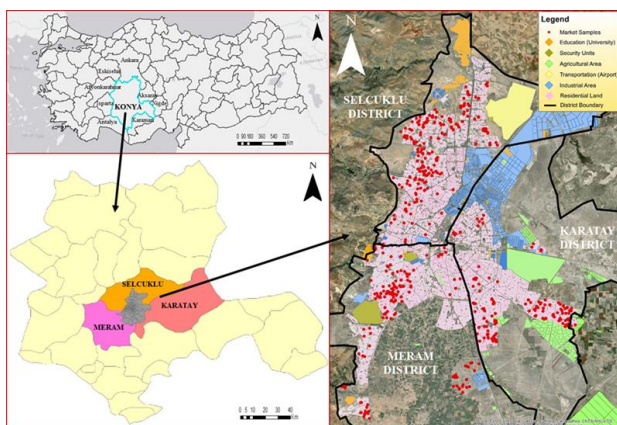


Figure 1. Location of the study region and distribution of sampling points

Generally, the spatial distribution of lands is outside the urban center. Because, in Konya's urban center, there was a problem in finding the land for sale, as the housing density was high. Thus, datasets with 32 features belonging to 319 market samples were obtained as follows (31 land feature and market value) (Table 1).

Feature	Min. Value	Max. Value	S _{dev}
Base area coefficient (BAC)	0.07	0.90	0.07
Floor area coefficient (FAC)	0.15	3.60	0.34
Number of floors (NF)	1.00	8.00	0.93
Building layout (BL)	1.00	3.00	0.35
Area of parcel (AP)	117	5462	639.20
Corner/Intermediate parc. (CIP)	1.00	2.00	0.48
Facade length (FL)	4.00	100	13.26
Facade number (FN)	1.00	4.00	0.62
Geometric shape (GS)	1.00	2.00	0.16
Infrastructure (IS)	0.25	1.00	0.21
Roadway type (RT)	1.00	2.00	0.48
Healthcare centers (HC)	1.00	10.00	2.01
Education Institution (EI)	1.90	10.00	1.92
Public agencies (PA)	1.00	10.00	1.67
Security units (SU)	1.00	10.00	2.57
Shopping centers (SC)	1.00	9.25	2.24
Cultural centers (CUL)	1.00	10.00	1.09
Entertainment centers (EC)	1.00	10.00	2.23
Green areas (GA)	3.00	10.00	1.42
Transportation (TR)	1.00	7.00	1.28
Insanitary areas (IA)	0.00	7.00	1.73
City Center (CTY)	0.00	5.00	0.65
Population density (PD)	0.75	233.08	33.20
Education level (EL)	0.87	2.32	0.28
Favourite neighborhood (FVN)	1.00	5.00	1.26
Residential density (RD)	0.35	146.67	12.92
Development potential (DP)	22.00	90.00	18.99
Geological condition (GC)	3.00	5.00	0.66
Climate condition (CC)	7.5	11.05	0.75
Air pollution (AIR)	5.9	7.0	0.14
Noise pollution (NOI)	1.0	4.85	1.04
Market value (₺ *1000)	55	2350	341.91

Table 1. Descriptive statistics of datasets

Features can have different types of continuous, categorical or non-scale variables as seen in Table 1. Besides, unscaled or non-pre-processing data can be misinterpreted in the model. Considering that such datasets may have negative effects on model analysis, Z-Score (Standard Score) transform was applied to features.

Z-Score (Standard Score) Transform: Z-score is simply the signed distance the variable value is from the mean in standard deviations (1). Z scores indicate how many standard deviations an observation is above or below the mean. These scores are a useful way of putting data from different sources onto the same scale.

$$z = \frac{x_i - \bar{x}}{\sigma_x} \quad (1)$$

z = z-score or standard score,

x_i = value of variables,

\bar{x} = arithmetic mean of variable values,

σ_x = standard deviation of variable values.

After the Z-Score transformation mentioned above, land values have been estimated with the SVR model. Thanks to the created SVR model, feature ranking/prioritizing has been realized according to the degree of effect in the land value estimation. The same process has been applied in the most used MRA method in real estate valuation and results have been compared.

2.2. Support vector regression (SVR)

SVR, the most comprehensive research of which has been carried out by Schölkopf and Smola (2004), is the regression form of SVM and is used in solving regression problems. In SVR, the main point is to minimize the generalized error limit instead of minimizing the error in the observed training cluster (Palaniswami and Shilton, 2002). It can use in applications such as linear/nonlinear regression or estimation of the unknown continuous function from noisy data. In regression, the general expression of support vectors is given below (2).

$$(X_1, Y_1), \dots, (X_n, Y_n); X \in R^d; Y \in R \quad f(X) = w \cdot X_i + b \quad (2)$$

Where; X_n = input in d-dimensional space, Y_n = outputs corresponding to the input vectors, w represents the weight vector and b represents the deviation. SVR can be implemented to regression issues by giving an alternative loss function (Awad and Khanna, 2015). SVR relied on defining a loss function called as an ϵ -insensitive (ϵ) loss function ignored errors that were within a certain distance of the true value (Wang et al., 2020). The main topic is to find the $f(X)$ function that can detect the Y_i value at a distance not greater than a previously estimated ϵ -insensitive (ϵ), using each X_i value (3).

$$f(X) = |Y_i - (w \cdot X_i + b)| \leq \epsilon \quad (3)$$

Algorithm finds that there is no error when it is less than ϵ , but it has an error when it is greater than ϵ . Support vectors perform linear regression in the high dimensional feature space using ϵ -function and, simultaneously, tries to eliminate model complicity by minimizing $\|w\|^2$ (Savas, 2019). However, linearity does not always occur. Therefore, the result can be found by transforming the input data into high-dimensional space using a kernel function in nonlinear applications.

Three control mechanisms affect the model estimation performance in the SVR method. These mechanisms are 1-Kernel type, 2-Kernel's function, and 3-Penalty/Regularization parameter (C). It is a selection of the most suitable one of linear, polynomial, sigmoid, radial based function (RBF) kernels of the model, expressed as the kernel type. Kernel function and regularization parameters are properties of the selected kernel type.

2.3. Multiple regression analysis (MRA)

MRA is a statistical analysis used to digitize the relationship between an output variable and multiple input variables. Regression analysis mainly aims to state the relationship between variables. The aim is to decide the contribution of each input variable to the total change in the output variable, and therefore to estimate the criterion value based on the value of the linear combination of the input variables. MRA method was used to test the SVR model results.

3. RESULTS AND DISCUSSION

3.1. Parameter selection from SVR model

In the study, models were established with different kernel types and parameters belonging to each type. Model results have been observed that the most successful one is the model established with the RBF function. The reason is that the RBF function can make specific judgments when input and output data are not linear. Two parameters (C and γ (gamma)) in the RBF kernel

absolutely affect the model estimation result. However, it is not known which parameter pair works best in the model. Therefore, it is necessary to determine the most suitable parameters in the relevant function. As a solution, selection of the most suitable C and γ (gamma) parameters for the RBF function was investigated according to the grid search method. In the first phase, C parameter has been tried by increasing its value by 1 in specified threshold intervals and the most suitable threshold value has been determined as " $C=2$ ". The effects of different C parameters on the model accuracy are shown over the R^2 value (Figure 2).

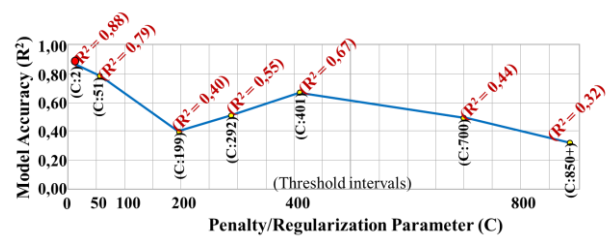


Figure 2. Model accuracies for different penalty/regularization (C) parameters

According to Figure 2, when C parameter is 2, the model has reached the highest R^2 value (0.88). When the model has the lowest R^2 value (0.32), the C parameter has determined as 851. In the second phase, a second search has been made in a smaller range near the parameters determined in the first phase with the grid search method and better parameter values have tried to be reached. However, in this part, not only C parameter but also γ value has been taken into account. C has tried with 0.1 increments between the threshold values 1-2. In addition, γ value has evaluated together with the C parameter in the range of 10^{-2} , 10^{-1} and 10^0 , respectively. The effects of (C , γ) pairwise parameters on the model accuracy are shown over the R^2 value (Figure 3).

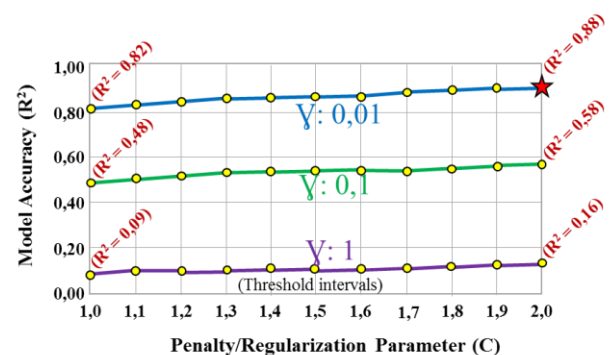


Figure 3. Model accuracies for different (C , γ) pairwise parameters

According to the results in Figure 3, the model has reached the highest R^2 value (0.88) when C is close to 2 and γ value is 10^{-2} . When the model has the lowest R^2 value (0.09), C and γ parameters have found as 1 and 10^0 , respectively. Consequently, the parameters in the RBF-SVR model for the land value estimation have determined as "2" for C and " 10^{-2} " for γ .

3.2. Model results

The estimation results of RBF-SVR and MRA models were compared with market values. The MRA method based on training data was selected to test both the SVR model and to examine whether the features that affect the value differ in different methods. Python v.3.8 was used for the SVR and SPSS

Statistics 22 was used for the MRA. When looking at the relationship between the estimation results obtained from the models and market value, it is seen that SVR gives slightly better results than the MRA (Figure 4-5). Although the successes for the two methods are very close to each other, it can be said that RBF-SVR model shows more effective results than MRA in the prediction of lands with lower market value.

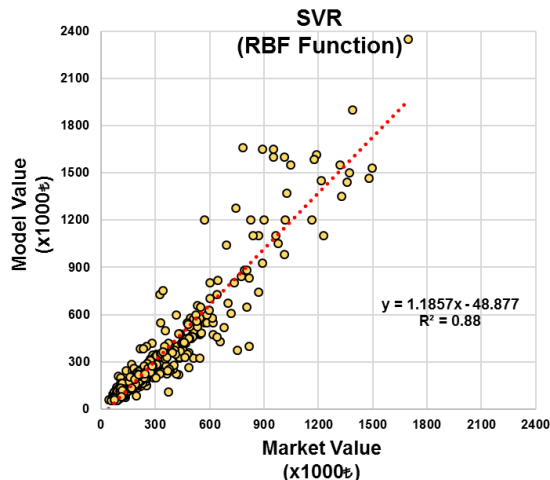


Figure 4. Comparison of SVR model with market values

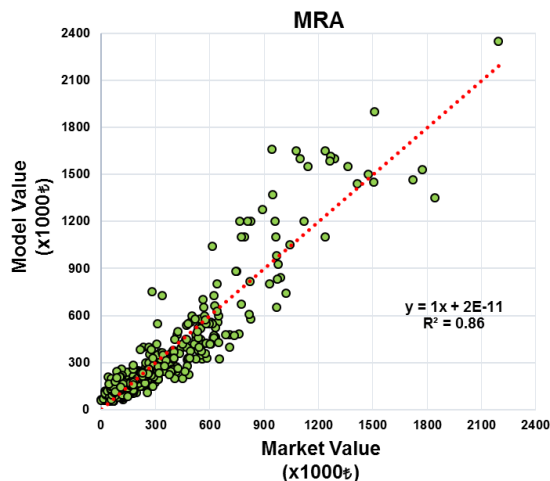


Figure 5. Comparison of MRA model with market values

After interpreting the compatibility of models with the market values, model performances have been examined by calculating MAPE (Mean Absolute Percentage Error), RMSE (Root Mean Square Error) and SD (Standard Deviation) (Table 2). Considering the results in table 2, SVR provided 0.13 RMSE while MRA gave 0.14 RMSE. SD values also gave close results. However, it should not be forgotten that even if the differences between the methods are small, they could correspond to serious differences in the value of real estate. Therefore, even if there are small nuance differences between the methods, the improvement of methods has very important in real estate estimation. Percentile differences can be seen from MAPE, which expresses the average ratio of absolute estimation errors to market values.

	RBF-SVR	MRA
R^2	0.88	0.86
SD	0.2710	0.3180
RMSE	0.1284	0.1366
MAPE%	20.7051	28.7152

Table 2. Model performance comparison of valuation models

Land value estimation results are based on the weight of features in the model. Therefore, the effect degree of input features in the model should be interpreted. This approach will help to produce a model with optimum features instead of the use of many features that affect the land value, and will also reveal meaningful relationships on the value. On the model estimate, effect degrees of features have examined by using SVR and MRA methods (Figure 6 and Figure 7). Through created models, feature ranking/prioritizing has been realized according to the degree of effect.

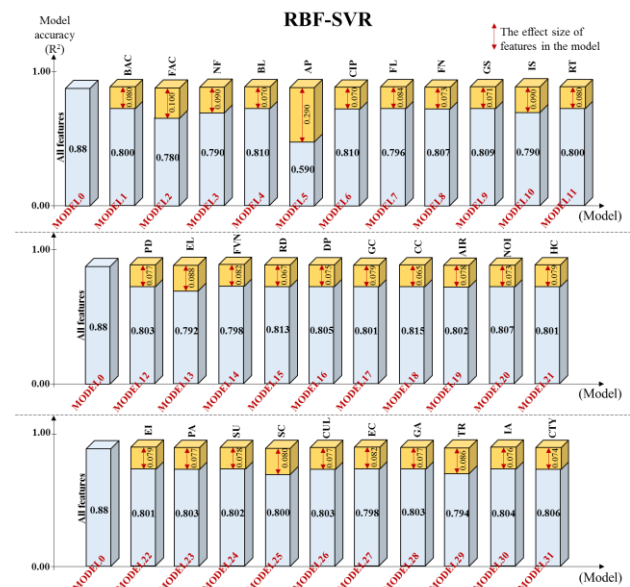


Figure 6. Effect degrees of features in the RBF-SVR model

According to findings (Figure 6), the most effective feature for the RBF-SVR model has been seen as Area of Parcel (AP). It has been found that the AP feature affected the accuracy of the model by approximately 0.29. It has the largest share of effect among all features. The most ineffective and also insignificant feature is Climate Condition (CC). The CC feature did not affect the success of the model much, and it had the least effect on the value. Generally, if the criteria for the RBF-SVR model are ranked in order of importance, from largest to smallest; AP > FAC > NF > IS > EL > TR > FL > FVN > EC > BAC > RT > SC > HC > EI > GC > SU > AIR > PD > PA > CUL > GA > IA > DP > CTY > NOI > FN > GS > BL > CIP > RD > CC ranked as. Legal features of the land stand out with high impact, while location-based and demographic criteria have less impact but very close impact. In the MRA method, regression coefficients have regarded for feature ranking/prioritizing (Figure 7).

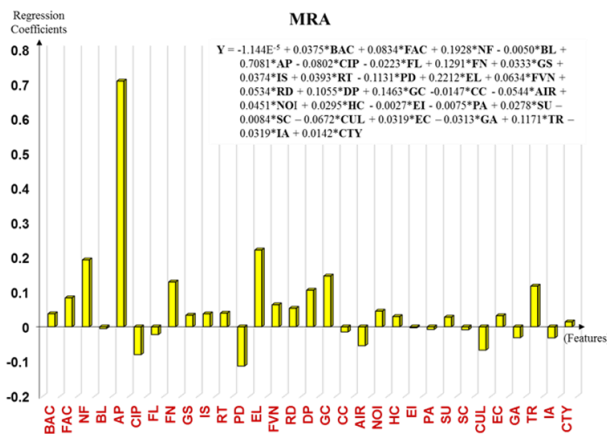


Figure 7. Effect degrees of features in the MRA model

According to findings (Figure 7), there are features that affect the land value as positive or negative. The most effective (positively) feature for the MRA model has been seen as Area of Parcel (AP). The weight coefficient of AP feature in regression was found as 0.7081. The lowest weight coefficient in terms of positive impact has been seen as proximity to the City center (CTY). On the other hand, there are also features that have a negative effect on value. Population Density (PD) appears to be the most effective feature in terms of negative impact. Education Institution (EI) appears to be the most ineffective feature in terms of negative impact. Generally, if the criteria for the RBF-SVR model are ranked in order of importance, from largest to smallest;

AP > EL > NF > GC > TR > DP > FAC > FVN > RD > NOI > RT > BAC > IS > GS > EC > HC > SU > CTY > EI > BL > PA > SC > CC > FL > GA > IA > AIR > CUL > CIP > PD ranked as. For both methods, the area of the parcel (AP), which is one of the features of the land, was determined as an effective factor.

4. CONCLUSION

By organizations (such as governments, municipalities) the objective, fair and reliable taxation about land, plot and real estate is crucial considering the sustainable land management, smart city applications and urbanization studies. However, real estate (land) values are affected by many factors such as differences in the economic conditions of the country, variation of people needs due to the living conditions, and socio-cultural structure. Therefore, the high number of factors makes mathematical modeling difficult for the mass real estate valuation. In the paper, the success of machine-learning approach, which provide ease of use in complex variation solutions, is examined in determining the real estate value. Moreover, it has been shown whether the optimum criteria can be created for mass valuation by determining the degree of impact of the features that affect the real estate value.

Within the scope of this study, by using the SVR method from machine learning algorithm for real estate valuation, the predict value of real estate was determined and the effects of features on the value was investigated. MRA method, which is the most preferred in real estate valuation, was used to test the SVR model. As a result;

- The success of SVR and MRA methods in estimating the market value of lands has been researched. Research results revealed that the SVR model is a more successful land value estimation model compared to the MRA model.

- According to both models, effects of features on the market value have been examined. In the review result, the most effective feature has been Area of Parcel in both models.

- It has been observed that the use of different approaches in the real estate value estimation varies in the impact degree of the features that affect the value.

- Although increasing the number of features that affect the value is effective for model performance, it is important to determine the most appropriate features specific to that region in model building.

ACKNOWLEDGEMENTS

The authors would like to thank interviewers who made efforts and respondents who patiently completed the survey in the data collection phase of the survey. This study was supported by the Scientific and Technological Research Council of Turkey (TUBITAK in Turkish) with 115Y769 Projects Number.

REFERENCES

Awad, M., Khanna, R., 2015. Efficient learning machines: theories, concepts, and applications for engineers and system designers, *Springer nature*, p. 268.

Baldominos, A., Blanco, I., Moreno, A. J., Iturrarte, R., Bernárdez, Ó., Afonso, C., 2018. Identifying real estate opportunities using machine learning. *Applied sciences*, 8(11), 2321.

Chang, D., 1996. Applications of the extent analysis method on fuzzy AHP. *European Journal of Operational Research*, 95(3), 649-655.

Chen, J. H., Ong, C. F., Zheng, L., Hsu, S. C., 2017. Forecasting spatial dynamics of the housing market using support vector machine. *International Journal of Strategic Property Management*, 21(3), 273-283.

Del Giudice, V., De Paola, P., Cantisani, G. B., 2017. Rough Set Theory for real estate appraisals: An application to Directional District of Naples. *Buildings*, 7(1), 12.

Doumpos, M., Papastamos, D., Andritsos, D., Zopounidis, C., 2020. Developing automated valuation models for estimating property values: a comparison of global and locally weighted approaches. *Annals of Operations Research*, 1-19.

Eren, H. M. E., Tudes, T., Yomralioglu, T., 1999. Coğrafi Bilgi Sistemlerinde Raster Tekniği İle Kent Taşınmaz Değer Haritalarının Üretilmesi.

Guan, J., Levitan, A. S., 1997. Artificial neural network based assessment of residential real estate property prices: A case study. *In Accounting Forum*, 20 (3/4), pp. 311-326.

Guan, J., Zurada, J., Levitan, A., 2008. An adaptive neuro-fuzzy inference system based approach to real estate property assessment. *Journal of Real Estate Research*, 30(4), 395-422.

Król, D., Lasota, T., Nalepa, W., Trawiński, B., 2007, June. Fuzzy system model to assist with real estate appraisals. *In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. pp. 260-269. Springer, Berlin, Heidelberg.

Lee, T. W., Chen, K., 2016. Prediction of House Unit Price in Taipei City Using Support Vector Regression. *APIEMS2016*, 12(7).

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

Palaniswami, M., Shilton, A., 2002, November. Adaptive support vector machines for regression. In *Proceedings of the 9th International Conference on Neural Information Processing*, 2002. ICONIP'02. (Vol. 2, pp. 1043-1049). IEEE.

Peris-Ortiz, M., Bennett, D. R., Yábar, D. P. B., 2017. Sustainable Smart Cities. *Innovation, Technology, and Knowledge Management*. Cham: Springer International Publishing Switzerland.

Peter, N. J., Okagbue, H. I., Obasi, E. C., Akinola, A. O., 2020. Review on the Application of Artificial Neural Networks in Real Estate Valuation. *International Journal*, 9(3).

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

Renigier-Bilozor, M., Biłozor, A., Wisniewski, R., 2017. Rating engineering of real estate markets as the condition of urban areas assessment. *Land Use Policy*, 61, 511-525.

Savas, B., 2019. Usage of Machine Learning Algorithms in Housing Value Estimation: Ankara Golbasi Application, *Konya Technical University, Institute of Graduate Studies, MSc Thesis*, Konya.

Schölkopf, B., Smola, A. J., 2002. Learning with Kernels – Support Vector Machines, *Regularization, Optimization and Beyond*. Cambridge, MA: The MIT Press.

Taffese, W. Z., 2007. February. Case-based reasoning and neural networks for real estate valuation. In *Artificial Intelligence and Applications*, pp. 98-104.

Unel, F. B., Yalpir, S., 2019. Reduction of mass appraisal criteria with principal component analysis and integration to GIS. *International journal of engineering and geosciences*, 4(3), 94-105.

Wang, X. J., Zeng, G. T., Zhang, K. X., Chu, H. B., Chen, Z. S., 2020. Urban Real Estate Market Early Warning Based on Support Vector Machine: A Case Study of Beijing. *International Journal of Computational Intelligence Systems*, 13(1), 153-166.

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.

Yalpir, S., Ozkan, G., 2018. Knowledge-based FIS and ANFIS models development and comparison for residential real estate valuation. *International Journal of Strategic Property Management*, 22(2), 110-118.