Gümüşhane University Journal of Science

GUFBD / *GUJS* (2024) 14(1): 116-130 doi: 10.17714/gumusfenbil.1363531 Research Article

Comparison of tree-based machine learning algorithms in price prediction of residential real estate

Mesken nitelikli gayrimenkul fiyat tahmininde ağaç tabanlı makine öğrenmesi algoritmalarının karşılaştırılması

Ayşe YAVUZ ÖZALP*[1], Halil AKINCI[1]

[1]*Artvin Çoruh University, Faculty of Engineering, Department of Geomatics Engineering, 08100, Artvin*

• Received: 20.09.2023 • Accepted: 31.10.2023

Abstract

Residential real estate is regarded as a safe and profitable investment tool while also meeting the basic human right to housing. The fact that there exists a large number of parameters both affecting the value of a house and varying based on place, person, and time makes the valuation process difficult. In this regard, accurate and realistic price prediction is critical for all stakeholders, particularly purchasers. Machine learning algorithms as an alternative to classical mathematical modeling methods offer great prospects for boosting the efficacy and success rate of price estimating models. Therefore, the purpose of this study is to investigate the applicability and prediction performance of the tree-based ML algorithms -Random Forest (RF), Gradient Boosting Machine (GBM), AdaBoost, and Extreme Gradient Boosting (XGBoost)- in house valuation for Artvin City Center. As a result of the study, the XGBoost and RF algorithms performed the best in estimating house value (0.705 and 0.701, respectively) as determined by the Correlation Coefficients (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) metrics. Thus, it can be said that ML algorithms, particularly XGBoost and RF, perform satisfactorily in residential real estate appraisal even with modest amounts of data and that the success rate grows as the amount of data increases.

Keywords: AdaBoost, GBM, RF, Residential real estate, Valuation, XGBoost

Öz

Mesken nitelikli gayrimenkuller, güvenli ve karlı bir yatırım aracı olarak kabul edilirken aynı zamanda temel insan hakkı olan barınma ihtiyacını da karşılamaktadır. Konutun hem değerini etkileyen hem de yere, kişiye ve zamana göre değişen çok sayıda parametrenin varlığı değerleme sürecini zorlaştırmaktadır. Bu bakımdan, doğru ve gerçekçi fiyat tahmini başta alıcılar olmak üzere sektörün tüm paydaşları için büyük önem taşımaktadır. Klasik matematiksel modelleme yöntemlerine alternatif olan makine öğrenme algoritmaları, fiyat tahmin modellerinin etkenliğini ve başarısını artırma bağlamında önemli olanaklar sunmaktadır. Dolayısıyla bu çalışmanın amacı, Artvin Kent Merkezinde ağaç temelli makine öğrenme algoritmalarından Rastgele Orman (RF), Gradyan Artırma Makinaları (GBM), AdaBoost ve Aşırı Gradyan Artırma (XGBoost) yöntemlerinin konut değerlemede uygulanabilirlikleri ve tahmin performanslarının araştırılmasıdır. Çalışmanın sonucunda, XGBoost ve RF algoritmaları, konut değerini tahmin etmede Korelasyon Katsayısı (Correlation Coefficients- R^2), Ortalama Mutlak Hata (Mean Absolute Error - MAE) ve Karesel Ortalama hata (Root Mean Squared Error- RMSE) ölçütlerinin tümünde (sırasıyla 0.705 ve 0.701) en iyi performansı göstermiştir. Böylece, başta XGBoost ve RF olmak üzere makine öğrenme algoritmalarının konut değerlemede az veri durumunda bile yeterli performans gösterdiği, veri setinin genişletilmesiyle birlikte başarı performansının daha da artacağı söylenebilir.

***Anahtar kelimeler**: AdaBoost, GBM, RF, Mesken nitelikli gayrimenkul, Değerleme, XGBoost.*

* Ayşe YAVUZ ÖZALP; ayavuzozalp@artvin.edu.tr

1. Introduction

Houses fulfill the fundamental human right to shelter and are recognized as a secure and highly profitable investment instrument. The importance of housing in the Turkish economy has grown further due to the facts that; (i) more than half of households in Türkiye own their home (Başer & Bozoğlu, 2019), (ii) a large portion of the private sector's wealth has been invested in the housing sector, (iii) the growth in the construction sector is greater than that in other sectors (Afşar & Yüksel, 2022), and (iv) real estate is regarded as an investment instrument to protect wealth against high inflation and a currency depreciation (Aydemir et al., 2020). On the other hand, population growth and changes in the demographic structure of society are predicted to raise housing demand by 1.65% by 2025 (Aydemir et al., 2020; Arslan et al., 2022). Furthermore, the increased demand for housing by foreigners in Türkiye in recent years (a 43.5% increase in 2021, according to the Turkish Statistical Institute) has made housing market research more significant and required.

An equitable and accurate real estate valuation procedure, which is an important component of land management, allows the real estate market to blossom, property and ownership rights to be secured, and real estate-based practices to be successfully implemented (Iban, 2021). On the other hand, the lack of an absolute value for houses has raised the importance of fair value (Doğan et al., 2022). The existence of a vast number of parameters that determine the value of the real estate, and the fact that such parameters vary based on time (subject to change over time), person, and place, make the valuation process time-consuming, costly, and inaccurate. In this regard, a concurrent value of a large number of real estate (i.e., mass appraisal) is required in many applications such as expropriation, land readjustment, and taxation, and is critical to ensure that the appraisal is completed in less time and at a lower cost (İlhan & Öz, 2020). At this point, fast-evolving information technology and machine learning (ML) techniques are gaining traction as viable alternatives to traditional mathematical modeling methods used in the late twentieth century (Ceh et al., 2018). Furthermore, these technologies have produced significant possibilities for streamlining this procedure and boosting the accuracy of price estimations (Louati et al., 2021). As a result, examining the applicability and contributions of artificial intelligence approaches in real estate appraisal has become vital and critical.

As known, real estate valuation, particularly mass valuation, is a difficult and multidimensional process. On the other hand, a comprehensive database containing information on the condition and value of real estate in Türkiye, coupled with restricted access to available data, further complicates the process of conducting valuation studies. Moreover, the residents' ability to make "unrealistic declarations" to the state regarding the purchase/sale of real estate in the twenty-first century (Sevgen & Tanrivermis, 2020) makes monitoring the real estate market extremely difficult. With a Decree that went into effect in 2019, the General Directorate of Land Registry and Cadastre (GDLRC) has been tasked with determining the value of real estate using mass valuation methods, producing value maps, and establishing and managing a value information database (İlhan & Öz, 2020). When the Real Estate Valuation Project initiated as part of the Land Registry Cadastre Modernization Project by GDLRC is completed, significant efforts will be taken to address such issues. It is also believed that investigating the applicability of algorithms that have arisen over time and in parallel with technological development in the real estate valuation business, particularly in different regions with diverse local dynamics, will contribute to this process.

In this context, numerous studies have been conducted to explore house price estimation and the factors that influence house prices. Following an examination of such studies, it was discovered that traditional and statistical methods such as Hedonic Modeling (HPM), Multiple Regression Analysis (MRA), and Nominal valuation (Yavuz Ozalp & Akinci, 2017; Tabar et al., 2021) were previously used, whereas ML algorithms have been widely used in more recent studies. The primary distinction between ML algorithms and traditional methods is that the valuation model is designed using input data rather than a set of directives (Louati et al., 2021; Adetunji et al., 2022), and models are trained by implementing various algorithms with the available data set in a given region (Iban, 2021). It has been discovered that ML algorithms are used either alone or in combinations (two or more) in these studies, and the common goal of the models created is to predict value by investigating the effect of real estate characteristics on value (Bilgilioğlu & Yılmaz, 2021). Borst (1991) conducted the first study in the literature that applied ML algorithms to the subject of real estate appraisal, employing the Artificial Neural Networks (ANN) method. Other studies utilizing ML techniques such as Support Vector Machines (SVM), Decision Trees (DT), Gradient Boosting Machines (GBM), eXtreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), and Random Forest (RF) followed this one. For example, MRA and RF were employed by Ceh et al. (2018), Yılmazer & Kocaman (2020), and Aydinoglu et al. (2021) (Table 1). In addition, Ravikumar (2016) used RF, SVM, and ANN, while Zaki et al. (2022) applied HPM and XGBoost. Moreover, the algorithms of Linear Regression (LR), Nearest Neighbor Regression (NNR), RF, and XGBoost were utilized by Hjort et al. (2022) whereas Yazdani (2021) employed HPM, ANN, RF, and k-Nearest Neighbors (kNN) ML algorithms. Most of this research revealed that ANN outperformed regression analysis (Sevgen & Tanrivermis, 2020; Alkan et al., 2022), whereas RF outperformed ANN (Ravikumar, 2016; Sevgen & Tanrivermis, 2020; Yazdani, 2021). The literature also stated that tree-based models such as RF and gradient-boosted trees produced excellent outcomes (Baldominos et al., 2018; Hjort et al., 2022).

Table 1. List of some literature on housing price prediction using ML algorithms.

References	Used method	Study area	Data source	The number of parameters	The number of data	R^2
Ceh et al. (2018)	HPM, RF	Ljubljana/ Slovenia	Official institutions	35	7407	0.230 0.570
Hayrullahoğlu et al. (2018)	CLLS, SR, ANN	Ankara/ Turkey	Websites	8	163	0.780 0.780
Ulvi & Özkan (2019)	ANN Fuzzy Logic	Konya/Turkey	---------	8	200	0.789 0.686
Yılmazer & Kocaman (2020)	MRA, RF	Ankara/ Turkey	GDLRC, Licensed expert	37	1162	0.710 0.749
Iban (2021)	RF, GBM, AdaBoost	Melbourne/ Avustralya	Websites (Kaggle)	15	6858	0.817 0.984 0.211
Adetunji et al. (2022)	RF	Boston / USA	Dataset	14	506	0.900
Alkan et al. (2022)	kNN, RF, SVM	Antalya/ Turkey	Websites	17	200	0.662 0.696 0.735
Ozdemir et al. (2022)	Deep Learn. RF, Pol. Regres.	Sakarya/ Turkey	Websites	23	213	0.861 0.772 0.765
Zaki et al. (2022)	XGBoost HPM	Boston/USA	Websites (Kaggle)	14	506	0.841 0.420

Although studies in the literature confirm that the RF algorithm outperforms other methods in real estate value estimation (Wang & Wu, 2018; Afonso et al., 2019; Yazdani, 2021; Embaye et al., 2021), it is still critical to investigate the method's performance, particularly with small data sets. To the best of our knowledge, despite the common use of ML algorithms (particularly ANN) in the literature, there were no studies found comparing the performance of the tree-based ensemble learning methods of RF, GBM, AdaBoost, and XGBoost. This approach distinguishes the current study from others and serves as the primary rationale for this research. Furthermore, the notion that housing markets are spatial markets with varying prices across geographies despite sharing comparable characteristics (Başer & Bozoğlu, 2019) is thought to offer novelty to such studies dealing with real estate valuation. As such, the purpose of this study is to explore the applicability and prediction performance of four tree-based ML algorithms (RF, GBM, AdaBoost, and XGBoost) in the situation of Artvin City Centre with limited/small data.

2. Material and method

2.1. Study area

This study was carried out in the city center of Artvin, located in the northeastern part of Türkiye, on the border with Georgia. Artvin Province is heterogeneous in terms of geographical structure, topography, and urbanization. Moreover, it also has a significant number of residential areas submerged under water as a result of the dams built in recent years, forcing urban migration and limiting the housing market. The study area is 768.91 hectares in size and is located between 41°47'24" and 41°50'24" East Longitudes and 41°9'36" and 41°11'42" North Latitudes (Figure 1).

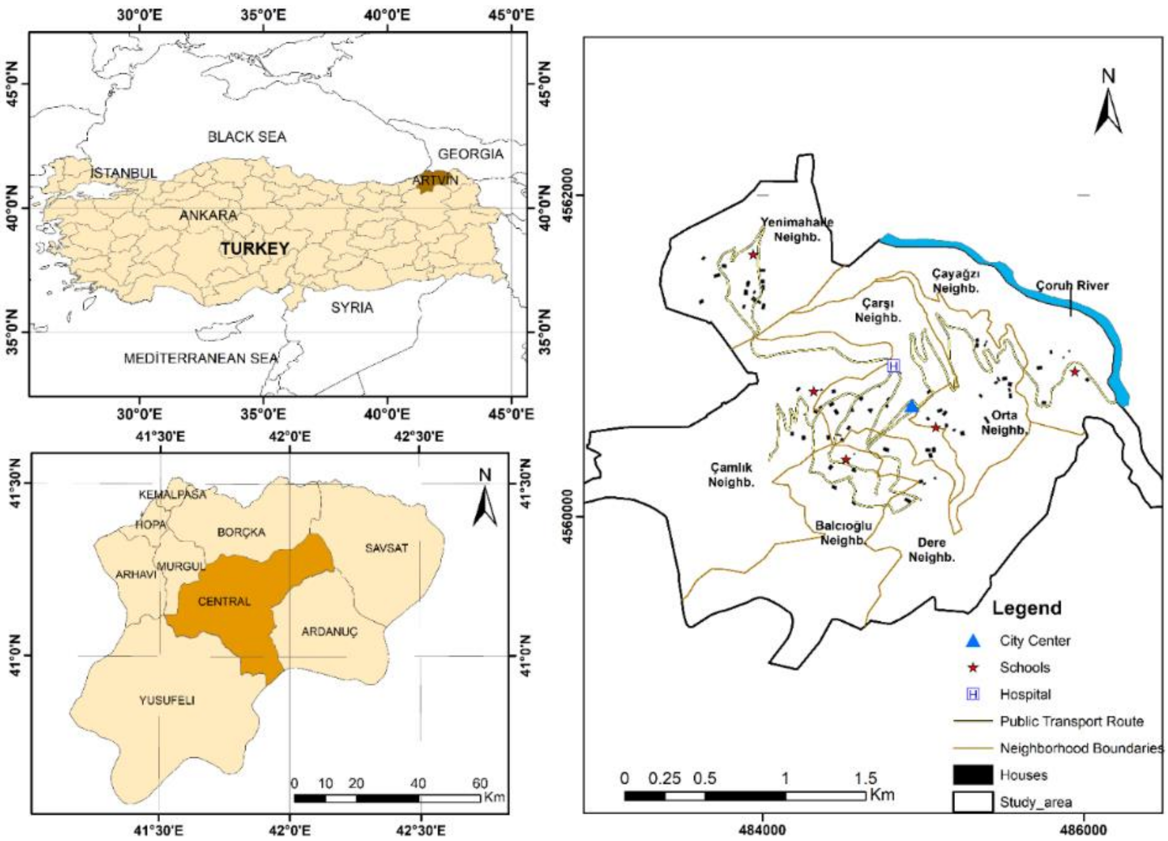

Figure 1. Study area map

The City Center, which is located on a hillside, has an elevation difference of 555 meters (rises from 195 to 750 meters), and the majority of the city center (approximately 80%) is above the 45% slope. A significant portion of the research region, which includes seven neighborhoods: Balcioglu, Camlik, Carsi, Cayagzi, Dere, Orta, and Yenimahalle, is geologically hazardous. Because of its difficult geography and environment, Artvin is one of the locations prone to landslides. According to Özalp et al. (2020), 68% of building blocks in the city center are in landslide zones with high and extremely high susceptibility. On the other side, the city's mountainous geography makes pre-construction work more expensive. In terms of land cover, Artvin is characterized by 90.89% forested and semi-natural areas, 6.47% agricultural zones (Corine, 2022), and has limited space dedicated to settlements.

On the other hand, five major dams -Muratlı (2005), Borçka (2006), Deriner (2012), Artvin (2015), and Yusufeli (2022)- were built on the Çoruh River flowing through Artvin province. The construction of these dams has led to the inundation, either in part or whole, of various settlements comprising 79 villages, neighborhoods, or districts, thereby leading to the loss of fertile agricultural terrain. Thus, 62.6 ha of residential area and 1960 ha of agricultural area submerged due to these dams (Yıldırımer et al., 2015). Migration to the city center from dam-evacuated villages, the opening of a university in the city in 2007, and the enactment of the Mortgage Act have all contributed to the dynamism of the Artvin real estate market (Table 2).

Table 2. Data on housing sales in Artvin Province and the Central District (TSI, 2022).

City/District	2015	2016	2017	2018	2019	2020	2021	2022
Artvin	1746	1825	1978	1783	2044	2075	1758	2060
Central district	513	445	483	378	482	409	354	342

When house sales in Artvin were analyzed in this regard, various numbers stood out, such as a 23.5% increase in January 2017 (142 houses), a 104.1% increase in July 2018 (200 houses), and the province with the highest rate of mortgaged sales in October 2019 (54.3%) (TSI, 2022). Table 2 highlights the strong sales volume in the central district in 2015. The reason for this is that the people whose real estate was expropriated owing to the aforementioned dams wish to purchase a home in the central district. The sales data presented in the table suggest that factors such as instability in the economy and increase/decrease in mortgage rates have an impact on the housing market. After all, Artvin was chosen as the research area to investigate the local dynamics of cities and their effects on valuation studies using scientific methodologies.

A total of 1758 residences were sold annually for the whole of Artvin province in 2021, with 619 of these being new houses. In addition, 354 of these house sales (109 of which were mortgaged transactions) occurred in Artvin's central district (Table 2). The total number of residences for sale was determined to be 204 when only the seven neighborhoods that formed the research area were examined (GDLRC, 2022).

2.2. Data set and application

The availability of the qualities that separate houses from one another (i.e., their quantitative and qualitative attributes) and the actual prices at which such houses were sold are required for reliable and accurate calculation of house prices. The most significant limitation of research on house value estimation is the lack of sufficient and reliable data.

It was discovered that the studies in the literature are highly diverse, particularly in terms of data collection. The price data used in the studies were obtained from sale advertisements posted on the internet (Hayrullahoğlu et al., 2018; Yılmazel et al., 2018; Aydemir et al., 2020; Alkan et al., 2022; Doğan et al., 2022), from the Central Bank of the Republic of Türkiye (Akay et al., 2019), and from valuation reports prepared by valuation experts that hold a license from Capital Markets Board (CMB) (Saraç, 2012; Esen & Tokgöz, 2021). The house values (actual sales prices) used in this study were gathered from real estate agents, CMB-licensed real estate valuation experts, home builders, and field studies that included face-to-face interviews with buyers/sellers. As a result, the data collection can be regarded to be trustworthy and accurately reflect market value. The method of data gathering distinguishes this study from others and assures that the study is conducted with actual sales prices.

Large data is generally preferred in the literature, particularly in studies examining ML methods; however, studies could also be carried out with small data, as evidenced by the studies by Büyük & Ünel (2021) with 73 houses, Saraç (2012) with 100 houses, Wilkowski & Budzynski (2006) with 114 houses, and Tabanoğlu (2019) with 150 houses. We were able to get the actual sales price of 103 out of 204 houses in Artvin City Center that were sold between January 1 and December 31 of 2021, as well as the attributes that were judged effective on their prices, using all of this information. As a result, the study's material comprises 103 residences with known actual sales prices.

Academic studies, international and national standards (IAAO, 2022; TDUB, 2022), and the geographical structure, physical qualities, and local dynamics of Artvin City Center were all taken into account in the selection of criteria that determine the value of houses. As a result, 17 parameters were determined by considering both the structural (gross area, age, number of rooms, floor level, number of balconies/bathrooms, elevator, parking lot, heating, aspect, and condition) and the spatial characteristics (distance to primary school/ hospital/ city center/ public transportation, Carsi Neighborhood, avenue/street). Data on the structural characteristics of the houses were obtained from buyers and sellers, from Building Permit Documents, and through field surveys. The data on spatial characteristics were obtained by measuring the actual walking distances with the help of ArcGIS 10.2 software on the current map obtained from the Municipality. The idea that better estimates can be made by considering spatial and non-spatial parameters together in the modeling process was also expressed by Wang & Wu (2018).

Descriptive statistical information on the numeric parameters of the selected 103 houses is presented in Table 3. The data in Table 3 give a general idea about the characteristics of the houses used in the study. Upon scrutinizing the data, it's apparent that the houses' ages span from 0 to 42 years. Among the 103 houses, 14 are newly constructed, whereas 11 have aged beyond two decades. Additionally, when assessing the properties in terms of size, it's clear that the residential areas span from 47 m^2 to 285 m^2. Notably, 18 of these properties encompass an area less than 100 m^2, while 20 exceed 200 m^2 in size.

Table 3. The numeric parameters in the data set with summary statistics

Parameters	Mean	Range (Min. - Max.)	Median
Sales price (TRY)	445790	110000- 920000	450000
Gross area (m^2)	127.75	47 - 285	130
Age	9.05	0 - 42	7
Floor level	3.95	-2 - 10	4
Number of rooms	2.83	1 - 6	3
Number of balconies	1.5	0 - 3	2
Number of bathrooms	1.43	1 - 3	1
Distance to primary school (m)	381.64	20 - 870	308
Distance to hospital (m)	1500.30	40 - 4355	1050
Distance to the city center (m)	1894.01	180 - 3950	1955
Distance to public transport (m)	105.48	3 - 585	35

As is well known, when using ML algorithms to model, the data set is divided into two parts: training data and test data. According to Sevgen & Tanrivermis (2020), the ratio of training sets could affect the result, and the ratio employed in the literature ranges between 60% and 80%. Indeed, training data ratios of 67%, 70%, 75%, 80%, and 90% were chosen by Wilkowski & Budzynski (2006), Aydinoglu et al. (2021), Hjort et al. (2022), Louati et al. (2021), and Akay et al. (2019), respectively. As a result, in this study, out of all data, 60% was randomly selected to be used for training whereas 40% was used for validation.

The prediction models were created using the scikit-learn ML toolkit for the Python programming language. In the modeling phase, the Probability Values (P-values) for all 17 parameters presumed to affect house values were computed. The P-value represents the probability of a relationship or difference arising between parameters in a data set randomly, even when there is no actual relationship or difference in reality (Ozdemir et al., 2022). As a result, the P-values of the parameters were checked one at a time and the parameters with the highest P-values were eliminated from the model and the P-values were recalculated. This procedure was performed until all parameter P-values were less than 0.05 (Table 4). Ultimately, only five out of 17 parameters were used in modeling according to the P-values.

Table 4. P-values of parameters used for RF, GBM, AdaBoost, and XGBoost algorithms

Parameters	XGBoost	GBM	AdaBoost	RF
Gross area	0.000	0.000	0.000	0.000
Age	0.004	0.011	0.001	0.002
Floor level	0.022	0.030	0.048	0.009
Aspect	0.000	0.001	0.000	0.000
Carsi neighborhood	0.001	0.011	0.000	0.003

2.3. Machine learning algorithms

Although there is various research in the literature that employ ML algorithms to predict house values (Yılmazer & Kocaman, 2020; Aydemir et al., 2020; Hong et al., 2020; Iban, 2021) (Table 1), there is no study that compares the performance of tree-based ensemble learning methods; RF, GBM, AdaBoost, and XGBoost. Therefore, the performances of these four approaches were compared in the current study.

2.3.1. Random Forest (RF)

RF, developed by Breiman (2001), is one of the supervised ML algorithms widely used in classification and regression problems. RF combines the output of multiple decision trees to produce a single result. In regression problems, the values produced by each tree are averaged, whereas in classification problems, the output of the RF model is determined according to the majority vote (Antipov & Pokryshevskaya, 2012; Chen et al., 2017; Avcı et al., 2023). The RF algorithm requires two hyperparameters to be set by the user: the number of trees (n_estimators) and the number of variables or tree depth (max_depth) (Hong et al., 2020). A Cross-validation technique is widely used in ML algorithms to obtain a more stable and reliable model prediction. In the case of RF, however, Ceh et al. (2018) reported that no cross-validation procedure was required to estimate model accuracy. In this study, the RF model was implemented in Python using the RandomForestRegressor function of the sklearn package. Following several attempts, the number of trees was fixed at 50 and the tree depth at four to ensure maximum accuracy in the application.

2.3.2. AdaBoost

Freund & Schapire (1997) coined the term AdaBoost, which stands for Adaptive Boosting. AdaBoost is the first practical boosting technique, designed to increase the performance of weak classifiers (Schapire, 2013). In general, boosting techniques attempt to construct a powerful classifier by merging numerous weak classifiers. AdaBoost's main principle is to train numerous weak classifiers for the same training set and then combine these weak classifiers to build a stronger classifier (Wu et al., 2020; He et al., 2021). The samples in the training dataset are given a weight value throughout the boosting procedure. Weights are then adjusted during iterations, with correctly categorized instances losing weight and wrongly classified instances gaining weight in the final iteration. The learning process is halted once the ideal weights are set in order to achieve the best performance (He et al., 2021). The AdaBoost model was implemented in Python in this work using the AdaBoostRegressor function from the sklearn package. In the model's implementation, the number of trees (n estimators) was set to 100, and the learning rate (learning rate) was set to 0.3.

2.3.3. Gradient Boosting Machines (GBM)

GBM, proposed by Friedman (2001), is a tree-based ensemble learning algorithm. The basic idea behind GBM is to bring together multiple weaker learners to improve their performance (Akinci, 2022). Unlike RF, GBM constantly constructs trees, with each tree attempting to repair the faults of the preceding tree (Wang et al., 2020). Gradient Boosting begins with the creation of the first leaf. The prediction mistakes are then used to form new trees. This process is repeated until the desired number of trees is obtained or the model can no longer be improved. The GBM model's hyperparameters are the number of trees (n estimators), the learning rate (learning rate), and the maximum depth of each tree (max depth) (Wang et al., 2020; Sahin, 2020). The GBM model was developed in Python in this work using the GradientBoostingRegressor function from the sklearn package. Following multiple trials, the number of trees (n estimators), and learning rate (learning rate) were set to 30 and 0.1, respectively.

2.3.4. Extreme Gradient Boosting (XGBoost)

XGBoost, developed by Chen & Guestrin (2016), is an efficient and optimized gradient tree-boosting algorithm. XGBoost, a type of supervised ML algorithm, is one of the most popular boosting algorithms that can be used for both regression and classification tasks. The most important factors behind the success of XGBoost are that it is scalable and runs ten times as fast as any available ML solutions (Chen & Guestrin, 2016). XGBoost uses a regularized boosting technique to prevent overfitting and to improve model accuracy (Can et al., 2021). In this study, the XGBoost algorithm was implemented using the Python XGBoost package (XGBoost, 2022). The XGBoost algorithm has five main hyperparameters: the number of decision trees (n_estimators), max depth of a tree (max_depth), learning rate (learning_rate), sample ratio of training data (subsample), and sample ratio of features (colsample_bytree). In this study, after several trials, the value of n_estimators parameter was taken as 50, max_depth as 3, learning rate as 0.1, and subsample and colsample_bytree as 1.

3. Results and discussion

3.1. Comparison of machine learning methods

As Gustafsson & Wogenius (2014) pointed out, developing a model that can identify the exact value of a house is quite challenging. The reasons for this are structural features, market conditions, the quality of materials used in the property, personal priorities/preferences of buyers and/or sellers, and problems in gathering and quantifying such information (Yavuz Ozalp & Akinci, 2017). Nonetheless, leveraging a comprehensive and diverse dataset, combined with the adoption of advanced machine learning algorithms, provides the potential to produce objective and realistic estimates of property values.

In this study, we analyzed the performance and usability of the models built with four machine learning algorithms, i.e., RF, GBM, AdaBoost, and XGBoost. The comparison of algorithm performance involved the use of fundamental metrics like the Correlation Coefficient (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) to evaluate the fit of regression models. The results can be found in Table 5. R^2 value, denoting the ratio of explainable variation to total variation and falling between 0 and 1, suggests a robust performance of the model if it is close to 1 (Yılmazer & Kocaman, 2020; Oral et al., 2021). In this context, considering the R^2 performance measure, which expresses the correlation between the actual sales values and the values estimated by the models, the models created with XGBoost and RF algorithms gave nearly identical results for this data set (0.705 and 0.701 respectively) and made more successful price predictions than AdaBoost and GBM (Table 5). Considering the training dataset, all algorithms produced successful results; however, the XGBoost and GBM methods outperformed RF and AdaBoost. The literature also confirmed that the accuracy of training data was higher than that of test data (Fei, 2020; Iban, 2021). Similarly, Yoshida & Seya (2021), and Hjort et al. (2022) also observed that the XGBoost algorithm consistently outperformed others in all predictions. In their study comparing ML algorithms including XGBoost and RF, Mete & Yomralıoğlu (2022) discovered that RF performed better than XGBoost.

Table 5. Performance measures for RF, GBM, AdaBoost, and XGBoost algorithms

Metrics	**XGBoost**	**GBM**	**AdaBoost**	**RF**
Accuracy of training data	0.85	0.83	0.77	0.77
Accuracy of test data	0.70	0.68	0.63	0.70
R^2	0.705	0.676	0.625	0.701
MAE	63623.12	70174.71	71985.45	66925.17
RMSE	78966.28	82601.52	88964.64	79451.96

The RMSE, a measure of the average deviation of the estimates from the observed values, varies between 0 and ∞, and lower values indicate better performance (Yılmazer & Kocaman, 2020; Oral et al., 2021; Ozdemir et al., 2022). When the algorithms were compared in this regard, the prediction model utilizing the XGBoost algorithm performed the best, followed by the model employing RF. The MAE metric displays the absolute error between observed and predicted values, and as the MAE metric value approaches zero, the predictive power of the model increases (Yılmazel et al., 2018). As shown in Table 5, the XGBoost algorithm delivered the best results yet again. These results imply that the methodologies can be used to value houses. However, the efficacy of the models in predicting house value was seen to be lower compared to some research in the literature (Table 1). The reason behind this is believed to be potentially linked to the limited sample size employed in model building. Similarly, Yılmazer & Kocaman (2020) linked the dependability of ML techniques to data amount and quality. From this perspective, the findings of this investigation based on real sales data are considered adequate.

3.2. Evaluation of parameters affecting house value

Although many structural and spatial parameters were taken into account, only five parameters were found to have an impact on the price. By applying four algorithms, the study unveiled that parameters like Gross area, Floor level, Carsi Neighborhood, Aspect, and Age significantly influence the price; however, the magnitude and significance of their effect and significance levels differ (Figure 2).

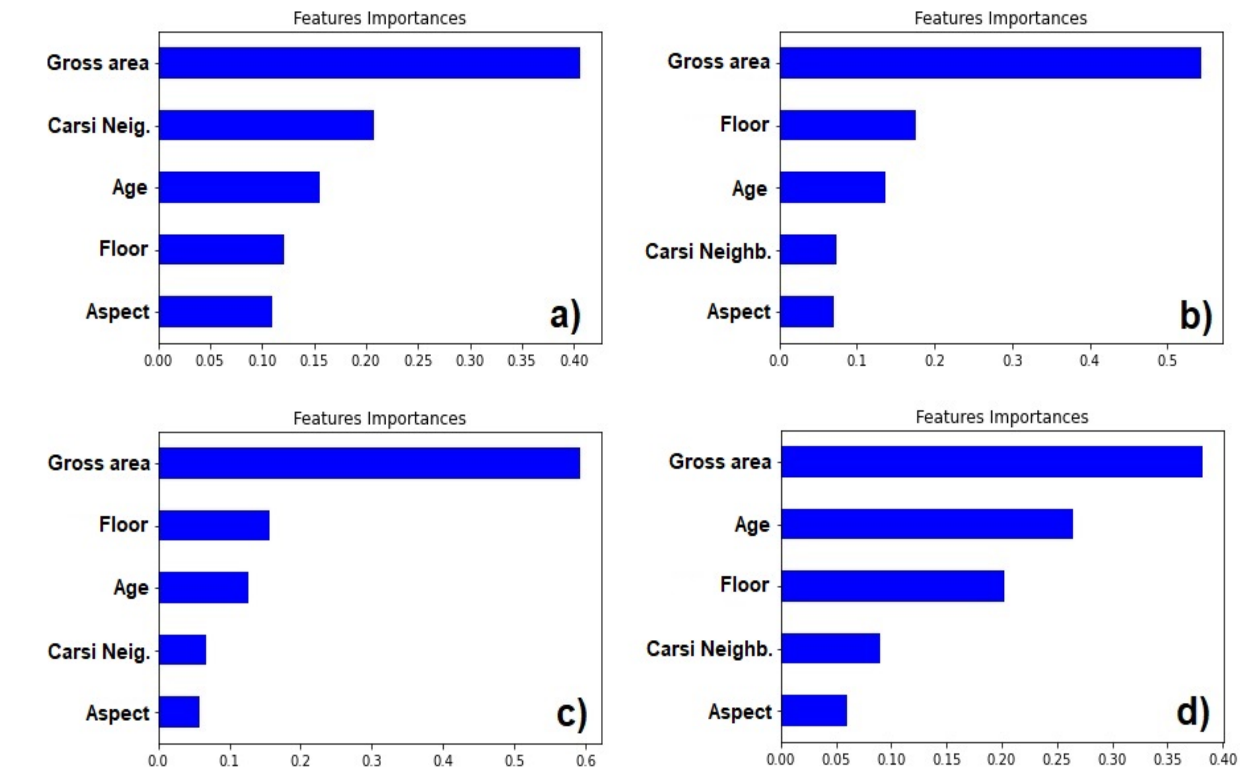

Figure 2. Features importance of ML algorithms a) XGBoost, b) GBM, c) RF, d) AdaBoost

When analyzing the significance levels of the parameters used in the estimation models, it becomes evident that the most crucial parameter across all models is the Gross area. Similarly, Yavuz Ozalp & Akinci (2017) discovered that the area parameter alone could explain 76% of the change in pricing in a study conducted on the basis of a neighborhood in Artvin. In another study conducted in the same area by Yavuz Özalp & Akıncı (2018) using 2015 sales data and the HPM approach, the parameters of floor area, age, floor level, and Carsi Neighborhood were found to be effective. Conversely, in that study, it was found that the development level parameter had an impact, whereas the aspect parameter did not show a significant effect. This finding indicates that the area, age, floor of the property, and proximity to the city center all have a substantial impact on the prices in Artvin City's real estate market. Although studies on the house-price relationship were conducted in different locations and using various methods, the area, age, and floor parameters were always found to have a determinant effect on price in the majority of studies (Yu et al., 2007; Sevgen & Tanrivermis, 2020; Yazdani, 2021). Even though spatial data boosted the adjusted R^2 value of the model, Hayrullahoğlu et al. (2018) found that structural factors were more beneficial than spatial and environmental characteristics.

The investigation concluded that the Carsi Neighborhood parameter had an effect on the pricing and caused the price to rise (Figure 3) since the Carsi Neighborhood is Artvin City Center's oldest and most central neighborhood, including key public amenities and attractions. This parameter proved the idea that location was very important and influential on price, as noted by Adetunji et al. (2022).

Upon examining Figure 3 and Table 6, it becomes apparent that the Gross area, Floor level, and Carsi neighborhood criteria have a favorable effect on the real sales price, whilst the Aspect and Age parameters have a negative effect. According to the level of significance, the three most effective characteristics were Gross area, Floor level, and Age, with Aspect being the least effective. The fall in property value with advancing age is consistent with theoretical projections. However, the effect of the Aspect parameter being negative is opposite to the expectation. In other words, despite the fact that aspects that receive more sunlight (such as South, Southeast, and Southwest) are favored in house preference, it was discovered in the current study to decrease the price. This could be because the city is located on a hillside and the slope faces northeast, there is an elevation difference due to its topographical structure, the study area is a small city and thus the volume of the real estate market is limited, and the owner's initiative in establishing the sales price.

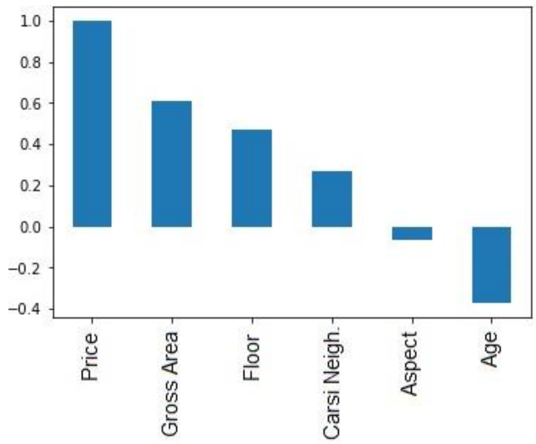

Figure 3. Correlation between price and other parameters

Table 6. Correlation values between sales prices and parameters

Parameters	Correlation
Price	1.000000
Gross area	0.609017
Floor level	0.472590
Carsi neighborhood	0.267000
Aspect	-0.065825
Age	-0.374230

Depending on the varied dynamics of different localities, the correlation of parameters with price, their significance, and positive/negative effects may emerge as contrary to expectations. Tuna et al. (2015) found, contrary to expectations, that the increase in the age of the building did not decrease the price of the house, and that older houses in central locations had higher prices. Iban (2021) found a correlation between the price and the number of bedrooms and bathrooms. Similarly, Oral et al. (2021) found that rental price, gross area of the house, and number of bathrooms were the most influential factors on the sale price. As a result, the fact that the price-parameter relationship of houses varies depending on location and local dynamics demonstrates the necessity and importance of conducting studies in various locations.

The actual sales values of the houses in the data set and the values predicted by ML algorithms were compared in the final stage of this study. According to the data set used, the prices of houses sold in Artvin City Center in 2021 range between 110,000 and 920,000 Turkish Liras (TRY) (Table 3). The same dataset shows that per-square-meter prices range from 1422.76 TRY to 9000 TRY, with an average of 3592.22 TRY. Using the test dataset and the prices predicted by the XGBoost algorithm, it was determined that the sales price in the test dataset ranged between 110,000 TRY and 700,000 TRY, with a minimum difference of 656 TRY and a maximum difference of 166,600 TRY between the actual sales price and the price estimated by the model (Figure 4). Between the estimated and real values, Sevgen & Tanrivermis (2020) found a minimum discrepancy of 600 TRY and a maximum difference of 60000 TRY. Oral et al. (2021) contended that the model performed better in estimating the prices of buildings with lower values.

When the graphs in Figure 4 are analyzed, it is clear that there are significant differences between the estimated prices and the actual prices of some houses, implying that houses are sometimes sold for much higher or much lower than the estimated prices. Again, the difference between expected and estimated price was greater for the GBM and AdaBoost algorithms, but less so for the XGBoost and RF algorithms. As a result of these findings, the XGBoost and RF algorithms can be considered suitable for application in mass valuation studies, with positive contributions in terms of time, labor, and cost.

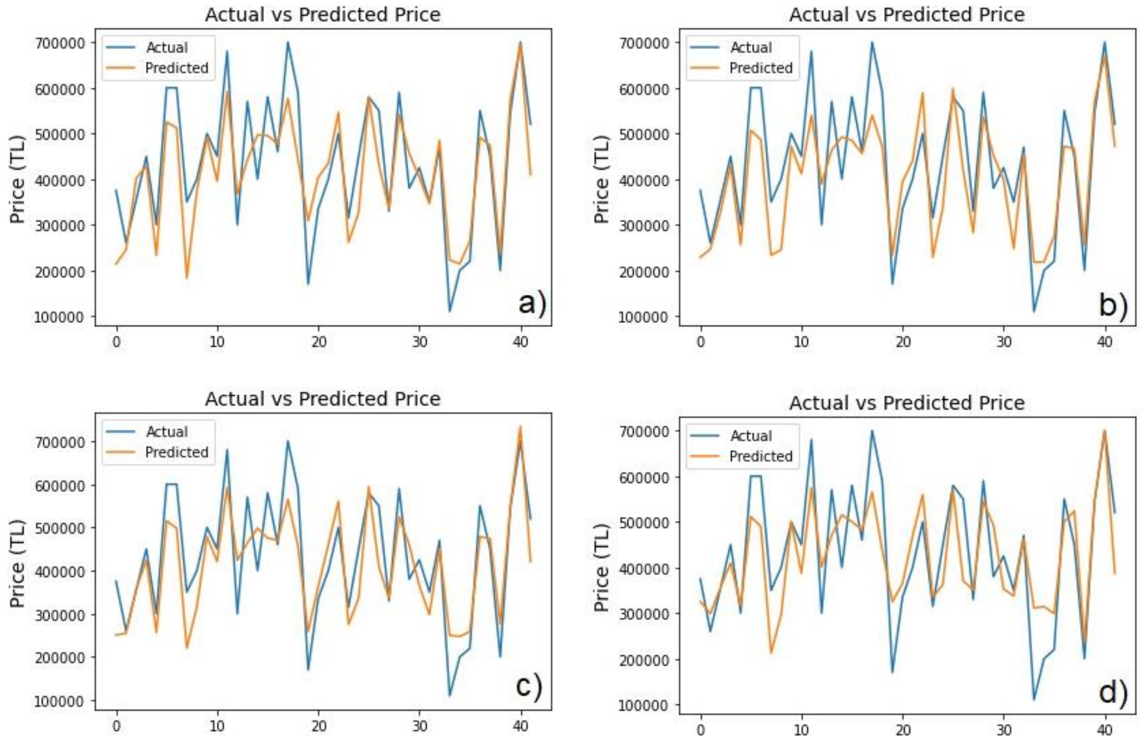

Figure 4. Actual versus predicted price for XGBoost (a), GBM (b), RF (c), and AdaBoost (d)

4. Conclusion

This study intends to investigate the performance of tree-based ML algorithms of Random Forest, Gradient Boosting Machine, AdaBoost, and XGBoost in estimating the value of houses and the influence of local dynamics on their prices for Artvin City Centre. According to the results, the XGBoost and RF algorithms performed the best (0.705 and 0.701, respectively) in assessing house value across all R^2, MAE, and RMSE measures. The relatively low success rate of the models developed in this study compared to their counterparts in the literature is due to the small number of samples used in model construction. However, it should be noted that ML methods, particularly XGBoost and RF, have demonstrated adequate performance in house valuation studies even with small or limited data and that the success rate would increase even further as the data set was expanded, implying that their use could save money, labor, and time. Besides, it is also very important to note that studies with sufficient accuracy can be carried out even if modeling has to be done with a small number of data, as in our country since the real sales prices cannot be recorded regularly. Because of the capacity of ML methods to mix multiple data sources, their precise, rapid, and objective predictive power, and their effectiveness even when the parameters are nonlinear, the application of ML algorithms -notably XGBoost and RF- should be considered as a reliable alternative in-house valuation. Moreover, it has been observed that the parameters including Gross area, Floor level, Carsi Neighborhood, Aspect, and Age significantly impacted the price according to all algorithms used in this study.

The scarcity of land suitable for construction due to its rugged topography, the costly pre-construction works, the city's remoteness from material supply centers, and the submergence of a significant portion of its residential districts due to the construction of several dams have all contributed to the dynamism of the Artvin housing market. As a result, the study's findings are meant to serve as a reference for all sector actors, including investors, homeowners, real estate brokers, and valuation specialists in Artvin's housing market.

Mechanisms for recording the actual sale prices of real estate, as well as the criteria used to establish their values, are still an ongoing process in Türkiye. In this regard, the GDLRC continues to determine and record the value of the real estate as part of the Land Registry Cadastre Modernization Project, resulting in the creation of a Value Information Database. As a result, it is critical that market price formation reflects the realistic and rational value of the real estate, and that regulations to ensure the correct handover price be specified in the declaration signed at the registry office be implemented as soon as possible, both in terms of establishing a successful housing policy and completing provision of taxes related to real estate sales. Finally, the application of RF and XGBoost algorithms in real estate valuation projects by GDLRC is thought to be advantageous by the benefits and as supported by the necessities indicated above.

Author contribution

A.Y.O.: Conceptualization, methodology, investigation, visualization, writing—original draft preparation, writing—review and editing. H.A.: Investigation, methodology, software, visualization, writing—original draft preparation, writing—review and editing.

Declaration of ethical code

The authors of this article declare that the materials and methods used in this study do not require ethical committee approval and/or legal-specific permission.

Conflicts of interest

The authors declare that there is no conflict of interest.

References

Adetunji, A.B., Akande, N., Ajala, F.A., Oyewo, O., Akande, Y.F., & Oluwadara, G. (2022). House price prediction using random forest machine learning technique. *Procedia Computer Science*, *199*, 806–813. https://doi.org/10.1016/j.procs.2022.01.100

Afonso, B.K.A., Melo, L.C., Oliveira, W.D.G., Sousa, S.B.S., & Berton, L. (2019). Housing prices prediction with a deep learning and random forest ensemble. *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2019)* (pp. 389-400), Salvador.

Afşar, M., & Yüksel, Ö.G. (2022). The effectiveness of the housing channel in monetary policy. *ESOGU Journal of Economics and Administrative Sciences, 17*(2), 345 – 367. https://doi.org/10.17153/oguiibf.1064471

Akay, E.C., Topal, K.H., Kizilarslan, S., & Bulbul, H. (2019). Forecasting of Turkish housing price index: ARIMA, random forest, ARIMA-random forest. *PressAcademia Procedia, 10*, 7-11. https://doi.org/10.17261/Pressacademia.2019.1134

Akinci, H. (2022). Assessment of rainfall-induced landslide susceptibility in Artvin, Turkey using machine learning techniques. *Journal of African Earth Sciences*, *191*, 104535. https://doi.org/10.1016/j.jafrearsci.2022.104535

Alkan, T., Dokuz, Y., Ecemiş, A., Bozdağ, A., & Durduran, S. (2022). Using machine learning algorithms for predicting real estate values in tourism centers. *Data Analytics and Machine Learning, 27*, 2601–2613. https://doi.org/10.1007/s00500-022-07579-7

Antipov, E.A., & Pokryshevskaya, E.B. (2012). Mass appraisal of residential apartments: An application of random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications, 39*, 1772-1778. https://doi.org/10.1016/j.eswa.2011.08.077

Arslan, Y., Ceritoğlu, E., & Kanık, B. (2022, October 14). *The effects of demographic changes on the long-term housing demand in Turkey.* Munich Personal Repec Archive. https://mpra.ub.uni-muenchen.de/52013/

Avcı, C., Budak, M., Yagmur, N., & Balcık, F. B. (2023). Comparison between random forest and support vector machine algorithms for LULC classification. *International Journal of Engineering and Geosciences, 8*(1), 01-10. https://doi.org/10.26833/ijeg.987605

Aydemir, E., Aktürk, C., & Yalçınkaya, M.A. (2020). Estimation of housing prices with artificial intelligence. *Turkish Studies, 15*(2), 183-194. http://dx.doi.org/10.29228/TurkishStudies.43161

Aydinoglu, A.C., Bovkir, R., & Colkesen, I. (2021). Implementing a mass valuation application on interoperable land valuation data model designed as an extension of the national GDI. *Survey Review, 53*, 349-365. https://doi.org/10.1080/00396265.2020.1771967

Baldominos, A., Blanco, I., Moreno, A., Iturrarte, R., Bernardez, O., & Afonso, C. (2018). Identifying real estate opportunities using machine learning. *Applied Sciences, 8*(11), 2321. https://doi.org/10.48550/arXiv.1809.04933

Başer, U., & Bozoğlu, M. (2019). Determination of the factors affecting housing rent using hedonic price model: the case of Ilkadım and Atakum districts of Samsun province. *Eurasian Journal of Researches in Social and Economics, 6*(4), 308-316.

Bilgilioğlu, S.S., & Yılmaz, H.M. (2021). Comparison of different machine learning models for mass appraisal of real estate. *Survey Review, 55*, 32-43. https://doi.org/10.1080/00396265.2021.1996799

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

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(l), 5–32.

Büyük, G., & Ünel, F. B. (2021). Comparison of modern methods using the python programming language in mass housing valuation. *Advanced Land Management, 1*(1), 23-31.

Can, R., Kocaman, S., & Gokceoglu, C. (2021). A comprehensive assessment of XGBoost algorithm for landslide susceptibility mapping in the upper basin of Ataturk Dam, Turkey. *Applied Science, 11*, 4993. https://doi.org/10.3390/app11114993

Ceh, M., Kilibarda, M., Lisec, A., & Bajat, B. (2018). Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. *ISPRS International Journal of Geo-Information, 7*(5), 168. https://doi.org/10.3390/ijgi7050168

Chen, T., & Guestrin, C. (2016). XG Boost: A scalable tree boosting system. *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD* (pp. 785–794), New York.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Tien, B., Duan, Z., & Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena, 151*, 147-160. https://doi.org/10.1016/j.catena.2016.11.032

Corine- Coordination of Information on the Environment. (2022, November 29). https://corinecbs.tarimorman.gov.tr/corine

Doğan, O., Bande, N., Genç Y., & Akyön, F.Ç. (2022). Estimation of housing fair values using artificial neural networks method in Kecioren/Ankara. *International Journal of Economic and Administrative Studies, 35*, 113-128. https://doi.org/10.18092/ulikidince.941952

Embaye, W.T., Zereyesus, Y.A., & Chen, B. (2021). Predicting the rental value of houses in household surveys in Tanzania, Uganda and Malawi: Evaluations of hedonic pricing and machine learning approaches. *Plos One. 16*, 1-20. https://doi.org/10.1371/journal.pone.0244953

Esen, Y., & Tokgöz, H. (2021). A different perspective to real estate valuation with fuzzy logic modeling. *Journal of Engineering Sciences and Design, 9*(4), 1155-1165. https://doi.org/10.21923/jesd.876523

Fei, Y. (2020). *California rental price prediction using machine-learning algorithms* [Master's Thesis, University of California Center for Social Statistics].

Freund, Y., & Schapire, R.E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences, 55*(1), 119–139.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

GDLRC- General Directorate of Land Registry and Cadastre. (2022, September 10). https://parselsorgu.tkgm.gov.tr/

Gustafsson, A., & Wogenius, S. (2014). *Modelling apartment prices with the multiple linear regression model*. Royal Institute of Technology. https://www.diva-portal.org/smash/get/diva2:725045/FULLTEXT01.pdfMultiple

Hayrullahoğlu, G., Aliefendioğlu, Y., Tanrıvermiş, H., & Hayrullahoğlu, A.C. (2018). Estimation of the hedonic valuation model in housing markets: the case of Cukurambar region in Çankaya district of Ankara province. *Ecoforum, 1,* 1-9.

He, Q., Jiang, Z., Wang, M., & Liu, K. (2021). Landslide and wildfire susceptibility assessment in Southeast Asia using ensemble machine learning methods. *Remote Sensing, 13*(8), 1572. https://doi.org/10.3390/rs13081572

Hjort, A., Pensar, J., Scheel, I., & Sommervoll, D.E. (2022). House price prediction with gradient boosted trees under different loss functions. *Journal of Property Research, 39*(4), 338-364. https://doi.org/10.1080/09599916.2022.2070525

Hong, J., Choi, H., & Kim, W.S. (2020). A house price valuation based on the random forest approach: the mass appraisal of residential property in South Korea. *International Journal of Strategic Property Management, 24*(3), 140–152. https://doi.org/10.3846/ijspm.2020.11544

IAAO- International Association of Assessing Officers. (2022, April 14). https://www.iaao.org/media/standards/StandardOnMassAppraisal.pdf

Iban, M.C. (2021). Accuracy analysis of community algorithms in estimating the value of real estate. *1st International Artificial Intelligence and Data Science Congress* (pp. 226-232), İzmir.

İlhan, A.T., & Öz, N.S. (2020). Applicability of artificial neural networks on mass valuation of real estates: the case of Gölbaşı District. *Hacettepe University Journal of Social Sciences, 2*(2), 160-188.

Louati, A., Lahyani, R., Aldaej, A., Aldumaykhi, A., & Otai, S. (2021). Price forecasting for real estate using machine learning: A case study on Riyadh city. *Concurrency and Computation: Practice and Experience*, 1-16. https://doi.org/10.1002/cpe.6748

Mete, M.O., & Yomralıoğlu, T. (2022). GIS and machine learning based mass valuation of residential properties. *11th Turkish National Photogrammetry and Remote Sensing Union (TUFUAB) Technical Symposium* (pp. 1-5), Mersin.

Oral, M., Okatan, E., & Kırbaş, İ. (2021). A study on house price prediction using machine learning methods: the case of Madrid. *3rd International young researchers student congress* (pp. 263-272), Burdur.

Özalp, A.Y., Akıncı, H., & Temuçin Kılıçer, S. (2020). Analysis of parameters affecting value of real estates with land property in Artvin Case. *Geomatik, 5*(2), 100-111. https://doi.org/10.29128/geomatik.579401

Ozdemir, M., Yıldız, K., & Büyüktanır, B. (2022). Housing price estimation with deep learning: a case study of Sakarya Turkey. *BSEU Journal of Science, 9*(1), 138-151. https://doi.org/10.35193/bseufbd.998331

Ravikumar, A.S. (2016). *Real estate price prediction using machine learning* [Master's Thesis, National College of Ireland School of Computing].

Saraç, E. (2012). *Real estate appraisal with artificial neural networks method* [Master's Thesis, İstanbul Kültür University Institute of Science].

Sahin, E.K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Science, 2,*1308. https://doi.org/10.1007/s42452-020-3060-1

Schapire, R.E. (2013). *Explaining AdaBoost. In Empirical Inference* (Bernhard Schölkopf, Zhiyuan Luo, Vladimir Vovk Eds.). Springer Berlin.

Sevgen, S.C., & Tanrivermis, Y.A. (2020). Mass apprasial with a machine learning algorithm: random forest regression. *Journal of Information Technologies, 13*(3), 301-311. https://doi.org/10.17671/gazibtd.555784

Tabar, M.E., Başaran, A.C., & Şişman, Y. (2021). Housing valuation study in Tokat province with multiple regression and artificial neural networks. *Turkish Journal of Land Management, 3*(1), 01-07. https://doi.org/10.51765/tayod.832227

Tabanoğlu, M. (2019). *Estimating the market value of residential buildings with artificial neural networks method: Düzce sample* [Master's Thesis, Düzce University Graduate School of Natural Sciences].

TDUB-Türkiye Değerleme Uzmanları Birliği. (2022, April 15). https://tdub.org.tr/uploads/documents/1667807030_d7767c42b3a070c20179

TSI- The Turkish Statistical Institute. (2022, November 10). https://data.tuik.gov.tr/Kategori/GetKategori?p=Insaat-ve-Konut-116.

Tuna, M.F., Türk, T., & Kitapçı, O. (2015). House prices with the help of linear regression and GIS estimating: the example of Ankara. *TMMOB HKMO 15. Türkiye Harita Bilimsel ve Teknik Kurultayı* (pp. 1-5), Ankara.

Ulvi, C., & Özkan, G. (2019). Usability of artificial intelligence techniques at real estate valuation and comparison of the methods. *Journal of Geomatics, 4*(2), 134-140.

Wang, C., & Wu, H. (2018). A new machine learning approach to house price estimation. *New Trends in Mathematical Sciences, 6*(4),165-171. https://doi.org/10.20852/ntmsci.2018.327

Wang, Z., Liu, Q., & Liu, Y. (2020). Mapping landslide susceptibility using machine learning algorithms and GIS: A case study in Shexian County, Anhui Province, China. *Symmetry, 12*, 1954. https://doi.org/10.3390/sym12121954

Wilkowski, W., & Budzynski, T. (2006). Application of artificial neural networks for real estate valuation. *Shaping the Change XXIII FIG Congress* (pp. 1-12), Munich.

Wu, Y., Ke, Y., Chen, Z., Liang, S., Zhao, H., & Hong, H. (2020). Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. *Catena, 187*, 104396. https://doi.org/10.1016/j.catena.2019.104396

XGBoost- XGBoost Python Package. (2022, December 13). https://xgboost. readthedocs.io/en/stable/python/

Yavuz Ozalp, A., & Akinci, H. (2017). The use of hedonic pricing method to determine the parameters affecting residential real estate prices. *Arabian Journal of Geoscience, 10*, 535. https://doi.org/10.1007/s12517-017-3331-3

Yavuz Özalp, A., & Akıncı, H. (2018). Using hedonic pricing model to analyze parameters affecting residential real estate value in Artvin City Center. *FIG Congress 2018* (pp. 1-15), İstanbul.

Yazdani, M. (2021). *Machine learning, deep learning, and hedonic methods for real estate price prediction* [Master's Thesis, Colorado University Department of Economics].

Yıldırımer, S., Özalp, M., & Erdoğan Yüksel, E. (2015). Determining loss and degradation of lands as a result of large dam projects and associated road constructions within the Coruh River Watershed. *ACU Journal of Forestry Faculty, 6*(1), 1-17. https://doi.org/10.17474/acuofd.00766

Yılmazel, Ö., Afşar, E., & Yılmazel, S. (2018). Using artificial neural network method to predict housing prices. *International Journal of Economic and Administrative Studies, 20*, 285-300. https://doi.org/10.18092/ulikidince.341584

Yılmazer, S., & Kocaman, S. (2020). A mass appraisal assessment study using machine learning based on multiple regression and random forest. *Land Use Policy, 99,* 104889. https://doi.org/10.1016/j.landusepol.2020.104889

Yoshida, T., & Seya, H. (2021). Spatial prediction of apartment rent using regression-based and machine learning-based approaches with a large dataset. *The Journal of Real Estate Finance and Economics*, 1-39. https://doi.org/10.1007/s11146-022-09929-6

Yu, D., Wei, Y.D., & Wu, C. (2007). Modeling spatial dimensions of housing prices in Milwaukee, WI. *Environment and Planning B: Urban Analytics and City Science, 34*(6), 1085–1102. https://doi.org/10.1068/b32119

Zaki, J., Nayyar, A., Dalal, S., & Ali, Z.H. (2022). House price prediction using hedonic pricing model and machine learning techniques. *Concurrency and Computation: Practice and Experience, 34*, 1-15. https://doi.org/10.1002/cpe.7342