Volume 1 Issue 2 2023
Pages: 149-157
ISSN: 2980-3152

CURRENT TRENDS IN COMPUTING

http://ctc.karabuk.edu.tr

RESEARCH ARTICLE

COMPARING THE PRACTICAL DIFFERENCES BETWEEN DECISION TREE AND RANDOM FOREST ALGORITHMS IN ESTIMATING HOUSING PRICES

PINAR SARI[1*], İPEK DOĞA BEDİRHAN[2] and ÇAĞRI SEL[3]

[1]*Department of Industrial Engineering, Karabük University, 78050, Karabük, Türkiye*
[2]*Department of Industrial Engineering, İzmir Democracy University, 35140, İzmir, Türkiye*
[3]*Department of IndustrialEngineering, Karabük University, 78050, Karabük, Türkiye*

ABSTRACT: The process of estimating the price of houses is becoming increasingly important in light of the changing economy worldwide, as houses are considered a basic need and a source of investment. This process is aimed at preventing losses, monitoring the market, minimizing problems, and arriving at accurate conclusions in the face of complex structures and issues. To achieve this, modern technology introduces the concepts of artificial intelligence and machine learning, which are integrated into all areas of life, to make progress in the process. Although machine learning and the algorithms used in this field have become widespread in recent years, there are still not enough studies on housing pricing. At the same time, people remain unaware of the field of machine learning and its applicability in every sector. Machine learning, in general, expands the data pool and enables new prediction results to be created by making future predictions based on data. The Decision Tree Algorithm, facilitating understanding and interpretation in every field, can handle multi-output problems and minimize preparation with its easy integration structure. The random forest algorithm can prevent overfitting problems in classification problems and can be applied in both regression and classification problems. The study aims to popularize the use of machine learning algorithms in the real estate sector. This will allow for effective housing price predictions during times of uncertainty and help in selecting the appropriate method by comparing the algorithms. Additionally, this study aims to reduce existing problems using these algorithms. A dataset called "California Housing Prices," containing 20.640 samples and eight features, was used in this study. The results of the Decision Tree and Random Forest Algorithms were examined in this dataset. Performance evaluation and comparison were made using MSE, RMSE, R2, and MAE metrics. It was observed that the Random Forest Algorithm produced better results and was superior to the Decision Tree Method when predicting house prices.

1. INTRODUCTION

Today's natural disasters, the pandemic period, actual-technological-psychological conflicts, supply-demand imbalance, ever-increasing individual or social expectations, etc., all effects it drives humanity to live each day for the next day. The new conditions bring with them economic uncertainties. In countries where the effects of uncertainty are felt intensely, this process is also reflected in the real estate sector. Houses, generally referred to as meeting basic shelter needs, are used as investments to create future security and as a means of protecting investment against the possible loss of value of money. In this case, predicting house values that have increased steadily over many years is essential. Sudden price changes overlooked and high-impact parameters, and frequently changing demand areas prevent buyers or sellers from finding the correct value. Price prediction applications should be used to prevent possible losses, make it easier to follow the market, and minimize problems during buying and selling [1].

E-mail address: 2328152501@ogrenci.karabuk.edu.tr (*)
Key words and phrases. Machine Learning, Decision Tree Algorithm, Random Forest Algorithm, Housing Price Estimation

The big data, complex structures, and unpredictable but frequent situations necessitate the integration of artificial intelligence into every aspect of life. Machine learning, which is a subset of artificial intelligence, is the key to transforming complex structures into accurate analyses.

This study examines the price estimation process of houses, which is of great importance in terms of need and investment, using machine learning algorithms and analyzing their performance [2]. The rest of the paper is organized as follows. Section 2 presents the literature review. Section 3 describes the decision tree method, the random forest algorithm, and MSE, RMSE, R^2, and MAE performance metrics. Section 4 declares the price estimations, the research findings, and various suggestions for future studies.

2.LITERATURE REVIEW

Machine learning is widely used in various fields beyond just housing price prediction. For instance, in Islamic banking and legal compliance, the k-NN algorithm is used to model daily price data and analyze lease certificate prices, attempting to predict future lease certificate prices [3]. In another study, demand forecasting is conducted for specific products in a local supermarket, accounting for various factors that affect product sales using electronic commerce data and machine learning techniques [4].

In the literature, some studies focus on estimating or classifying house prices, while others compare different methods and their performances. For instance, a recent study utilized machine learning algorithms to predict potential housing prices and compared the prediction performances [5]. The classification estimate of housing prices in Kütahya is obtained by running a model, which includes various criteria and data affecting prices, in machine learning classification algorithms [6]. House price forecasting in Sakarya determined the effect of parameters on house prices using deep learning and machine learning methods [7]. In Nevşehir, as a result of survey and site data, factors affecting the price of sold flats are determined by the Hedonic Price Method [8]. Another study on the Hedonic Price Method includes factor analysis evaluation with survey data conducted in Erzincan [9]. Additionally, a system has been developed that collects and records housing data for sale on the Internet and predicts housing prices using 14 different algorithms [10].

Several research studies have been conducted to compare the performance of different models in predicting the housing price index [11]. These studies involved making price predictions using various machine-learning algorithms and comparing the results [12]. Additionally, some studies focused on comparing the accuracy of ANN models and three different regression models used for house price prediction [13-15].

Some of the articles examined during the literature review and the methods they included are listed in Table 1. The purpose of creating the table is to clearly see the algorithms used in this study and other studies.

Machine learning is widely used in various fields beyond just housing price prediction. For instance, in Islamic banking and legal compliance, the k-NN algorithm is used to model daily price data and analyze lease certificate prices, attempting to predict future lease certificate prices [3]. In another study, demand forecasting is conducted for specific products in a local supermarket, accounting for various factors that affect product sales using electronic commerce data and machine learning techniques [4].

In the literature, some studies focus on estimating or classifying house prices, while others compare different methods and their performances. For instance, a recent study utilized machine learning algorithms to predict potential housing prices and compared the prediction performances [5]. The classification estimate of housing prices in Kütahya is obtained by running a model, which includes various criteria and data affecting prices, in machine learning classification algorithms [6]. House price forecasting in Sakarya determined the effect of parameters on house prices using deep learning and machine learning methods [7]. In Nevşehir, as a result of survey and site data, factors affecting the price of sold flats are determined by the Hedonic Price Method [8]. Another study on the Hedonic Price Method includes factor analysis evaluation with survey data conducted in Erzincan [9]. Additionally, a

system has been developed that collects and records housing data for sale on the Internet and predicts housing prices using 14 different algorithms [10].

Several research studies have been conducted to compare the performance of different models in predicting the housing price index [11]. These studies involved making price predictions using various machine-learning algorithms and comparing the results [12]. Additionally, some studies focused on comparing the accuracy of ANN models and three different regression models used for house price prediction [13-15].

Some of the articles examined during the literature review and the methods they included are listed in Table 1. The purpose of creating the table is to clearly see the algorithms used in this study and other studies.

TABLE 1. Techniques Used in Housing Price Estimation

Study	k-NN	ANN	Regression	NB	Decision Tree	Random Forest	Hedonic Pricing	ARIMA
[3]	X							
[4]		X	X					
[5]		X	X					
[6]	X			X	X	X		
[7]		X	X			X		
[8]							X	
[9]							X	
[10]						X		
[11]						X		X
[12]		X			X			
[13]		X						
[14]			X					
[15]			X					
This study					**X**	**X**		

As a result of the literature review conducted under the leadership of the relevant subject and its sub-headings, it has been observed that the subject has received more attention in recent years compared to previous years and contains methods that can be applied in many fields. However, in addition to the need for detailed research on the topic of house price forecasting, especially in national articles, it is thought that machine learning is overlooked in the field of artificial intelligence. This study aims to make machine learning accessible to more people and to be a new resource for comparing algorithms on house price prediction.

3. METHOD

This section presents the decision tree algorithm and the random forest algorithm proposed in the study. In this area, the concept of machine learning is mentioned in detail, and the Decision Tree and Random Forest Algorithms included in the study are explained.

3.1 Machine Learning

Machine learning, which Arthur Lee Samuel first used in designing a checkers game, enables making future predictions by obtaining inferences from existing data. Thus, more data leads to a larger pool of information, generating new predictions. Since the data has complex structures, the human factor is ineffective in the interpretation and development phase. Machine learning comes into play at this point, making it easier to analyze complex data accurately in the future.

Machine learning is a layer that exists between artificial intelligence and deep learning. Assuming that neural networks are the fundamental layer, the sequence of progression is neural networks, deep learning, machine learning and artificial intelligence. Similar to other algorithms, machine learning yields more accurate results as it has access to more data. Unlike traditional programming, the process of machine learning is automated with minimal human input of data.

In the past few years, machine learning has been categorized into three groups, namely supervised, unsupervised, and semi-supervised learning. In supervised learning, the system is trained using labeled data, which means that the training data is fed with manual labels to enable the system to learn the problems. An example of supervised learning can be a system that is trained to evaluate whether an email is spam or not.

Unsupervised learning is a type of machine learning where the system uses unlabeled data during training to discover patterns without being taught what the desired results are. The system groups data based on similarities and differences. For example, social media suggestion fields can be created using this technique [16].

Semi-supervised learning is a combination of labeled and unlabeled data, which provides better improvements. This method is useful in applications such as autonomous vehicles. Reinforcement learning involves the model gaining experience by taking actions in its environment and learning through reward and punishment. This approach is useful in developing online advertising programs or industrial/service robots by determining effective strategies based on feedback and results.

3.1.1 The Decision Tree Algorithm

It consists of decision nodes, leaf nodes and branches between them, according to the features and planned goal. It can be easily adapted to classification and regression problems and is easy to interpret and understand. It offers advantages such as low cost, visualizability, and a structure that provides easy integration, reliability, and minimization of preparation.

In the method, an objective function is defined to be optimized. However, the relevant notation (f: the attribute that will do the division, D_p: datasets belonging to the main node, D_j: datasets with the jth node, I: impurity criterion, N_p: total training data, N_j: number of samples in the jth node, m: the number of child nodes obtained as a result of division) and the following formula (1) is used.

$$IG\ (D_p, f) = I(D_p) - \sum_{j=1}^{m} \frac{N_j}{N_p}\ I(D_j) \qquad (1)$$

3.1.2 Random Forest Algorithm

Random Forest Algorithm, one of the widely used ensemble classification algorithms, was developed by Leo Breiman and Adele Cutler. In addition to classification and regression problems, it provides a repeatable and deep-dive structure in the cluster discovery process. In the method, there is a direct relationship between the number of trees and the result obtained because the increase in the number of trees means the result gets closer to accuracy. A decision forest is formed due to the combination of many decision trees. The difference between the random forest algorithm and the decision tree algorithm is that the steps of obtaining the root node and dividing the nodes work randomly in the random forest algorithm. Preparation of the dataset consists of creating a decision tree for each step and accordingly creating the estimated value of each decision tree, performing voting for each value resulting from the prediction, and creating a result by selecting the most voted prediction.

Today, the random forest algorithm is widely used to find loyal customers or to make credit risk assessments in the finance banking sector. It can perform disease definition, medical device analysis, gene analysis, and drug-component combinations in medicine. The method benefits from predicting stock behavior within the stock market and increasing customer satisfaction predictions in the e-commerce industry or similarly developing a customer recommendation system.

Random Forest is a solution approach for incomplete data. It prevents overfitting when enough trees are used. However, it can be time-consuming and not ideal for creating new data points outside the discrete set. It is possible to reduce the variance of an estimation by using the average of many estimates. The bagging process is calculated with Equation (2). Here, M is the number of different trees trained with many data subsets and selected by random replacement. fm is the whole unit, and m is the number of trees.

$$f(x) = \frac{1}{M} \sum_{m=1}^{M} f_m\ (x) \qquad (2)$$

Equation (3) is used in the random forest algorithm, where the Gini criterion, called the impurity measure that measures the homogeneity of the dataset in a node, is used. Here C_i: i. where T is a randomly selected sample in the training set, $f(C_i, T)$: represents the probability that the samplein the T set belongs to class C_i, and |T| is the number of samples in the T set [17].

$$\sum_i \sum_{j:i\neq j} \left(\frac{f(C_i,T)}{|\mathrm{T}|}\right)\left(\frac{f(C_j,T)}{|\mathrm{T}|}\right) \qquad (3)$$

4. EXPERIMENTAL STUDIES AND FINDINGS

In the study, the dataset named "California Housing Prices" taken from the *scikit-learn* library is analyzed. Decision Tree and Random Forest Algorithm results were examined on the data using the Python programming language. The dataset contains average income (MedInc), average age of houses (HouseAge), average number of rooms per household (AveRooms), average number of bedrooms per household (AveBedrms), block population (Population), average household occupancy (AveOccup), decimal degree features.

Using Feature Importances, how much importance the Random Forest Algorithm attaches to each feature in the dataset was measured. These values show the effect of the features in the dataset on prediction performance. If a feature has a higher feature importance value, it means that it plays a more significant role in prediction than other features. The importance of the features is listed in Figure 1. The feature that contributes the most to the prediction is the "MedInc" feature.

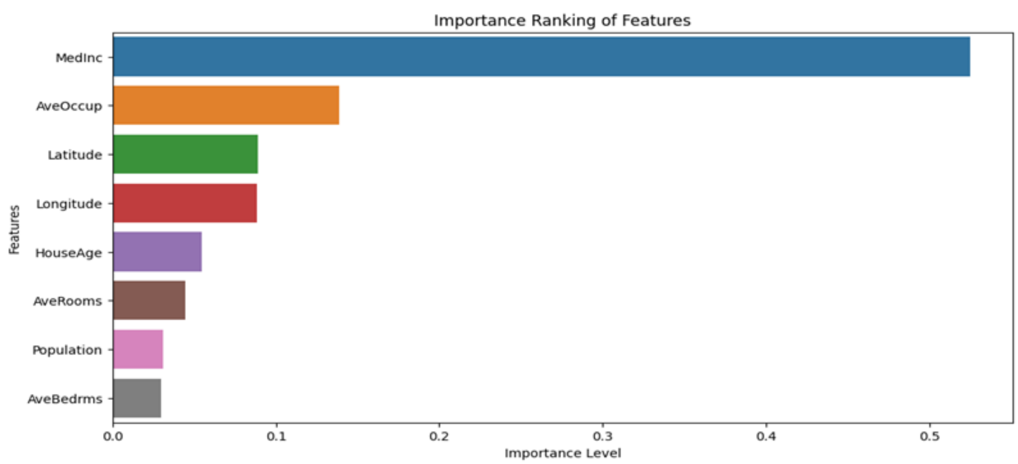

FIGURE 1. Importance Ranking of Features

The dataset has 20.640 examples, including latitude in decimal degrees and longitude in decimal degrees. By comparing the features, the horizontal column chart in Figure 1 was created according to the order of features' importance. The statistical values of the features in the dataset are as in Table 2. Histogram graphs of the features in the dataset are shown in Figure 2.

TABLE 2. Features and Values Table

Features	Mean	Standart Deviation	Min Value	%25	%50	%75	Max Value
MedInc	3.870671	1.8999822	0.499900	2.563400	3.538400	4.743250	15.00100
HouseAge	28.639486	12.585558	1.000000	18.00000	29.00000	37.00000	52.00000
AveRooms	5.429000	2.474173	0.846154	4.440716	5.229129	6.052381	141.90909
AveBedrms	1.096675	0.473911	0.333333	1.006079	1.048780	1.099526	34.066667
Population	1425.47674	1132.46212	3.000000	787.000000	1166.000	1725.000	35682.000
AveOccup	3.070655	10.386050	0.692308	2.429741	2.818116	3.282261	1243.3333
Latitude	35.631861	2.135952	32.54000	33.930000	34.26000	37.71000	41.950000
Longitude	-119.56970	2.003532	- 124.3500	-121.8000	-118.4900	-118.0100	-114.3100

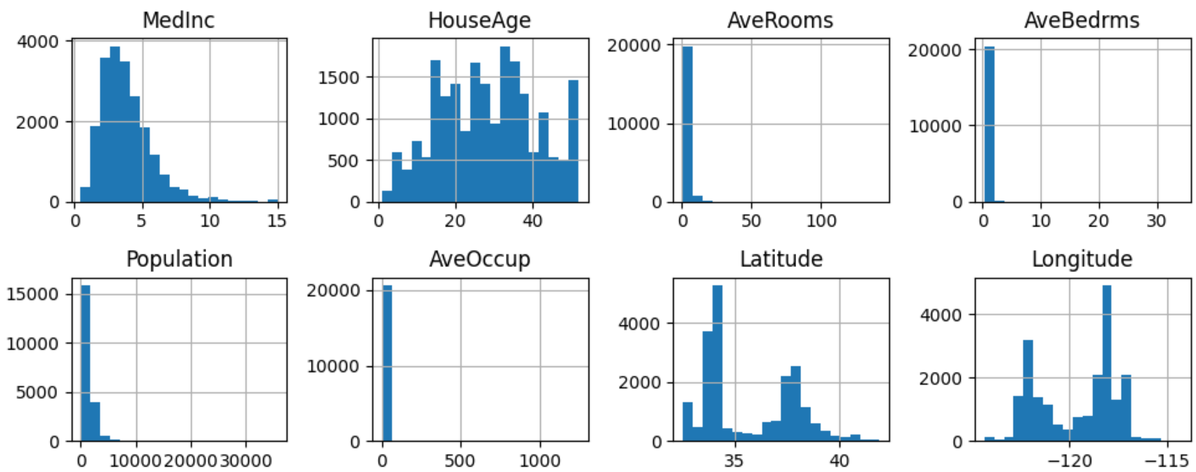

FIGURE 2. Histogram Charts Created According to the Distribution of Features

4.1.Performance Evaluation

Performance evaluation is a process to test a model or system's effectiveness, accuracy and efficiency. It contributes to determining closeness to the target and analysis of process effectiveness. Measurable metrics and data are generally used in the evaluation process.

4.1.1. Coefficient of Determination (R^2) / Adjusted R^2

R^2 (Coefficient of Determination) is an evaluation method for measuring the fit of a regression model, or in other words, how well it fits the observed data (Equation (4)). The value in the range 0-1 is usually expressed as a percentage. The R^2 value being close to 1 indicates that the model fits the data. Here, error variance shows how much the predicted values deviate from the actual values. The average dependent variable variance represents the overall variance of the dependent variable.

$$R^2 = 1 - \left(\frac{Error\ Variance}{Average\ Dependent\ Variable\ Variance}\right) \qquad (4)$$

If independent variables are added, the model has a more complex structure and the R^2 value increases accordingly. Complexity gives rise to 'overfitting'. The corrected R^2 method is used in cases where there are too many independent variables in the model or the desired success is not achieved in the test. *n* is the number of sample size and *p*indicates the number of independent variables in Equation 5 that is used tocalculate the corrected R^2.

$$Corrected\ R^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1} \qquad (5)$$

4.1.2.Mean Squared Error

Mean Squared Error (MSE) measures the distance between predicted and actual values. MSE shows the average square differences. MSE closer to 0 refers to a good estimation. MSE uses error squares; thus, significant errors are punished more than minor ones. RMSE and other performance measures can be used in case comparison with original units is difficult. MSE is calculated by Equation 6 where *n* is the number of observations, y_i is the real value and a_i is the prediction made by the model.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - a_i)^2 \qquad (6)$$

4.1.3.Root Mean Squared Error

Root Mean Squared Error (RMSE) is the square root version of MSE and provides a meaningful expression on the original measurement units. RMSE is easy to interpret due to its compatibility with the original units. Just like MSE, major errors are more affected than small errors. RMSE value close to 0 indicates that the model makes a good prediction. RMSE is calculated by Equation 7, where *n is* the number of observations, y_i is the real value and, a_i is the prediction made by the model.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - a_i)^2} \tag{7}$$

4.1.4.Mean Absolute Error

When calculating the Mean Absolute Error (MAE) value, the first thing to consider is the difference between the actual values and the estimated values for each observation in the dataset. The average of these differences is then calculated by Equation 8. It takes values between 0-∞ and the lower the value, the better performance is achieved. MAE value close to 0 indicates that the model makes a good prediction. The use of error values together with absolute values prevents large errors from being affected too much by small errors. However, some mathematical disadvantages may be encountered compared to other metrics. In Equation 9, n donates the number of observations, y_i is the real value, and a_i is the prediction made by the model.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - a_i| \tag{8}$$

Performance evaluation and comparison is presented using MSE, RMSE, R^2 and MAE metrics in Table 3. For the dataset examined, the fact that RMSE, MSE and MAE metrics are low and the R^2 value is close to one show that the estimation of the algorithm is good. RMSE, MSE and MAE values of Random Forest Algorithm are lower than the Decision Tree Algorithm, and the R^2 value of the Random Forest Algorithm is closer to one than the R^2 value found with the Decision Tree Algorithm.

TABLE 3. Comparison of Algorithms and Related Metric Values

Algorithm	MSE Value	RMSE Value	R^2 Value	MAE Value
Decision Tree	0.495235	0.703729	0.622076	0.454679
Random Forest	0.255368	0.505339	0.805123	0.327542

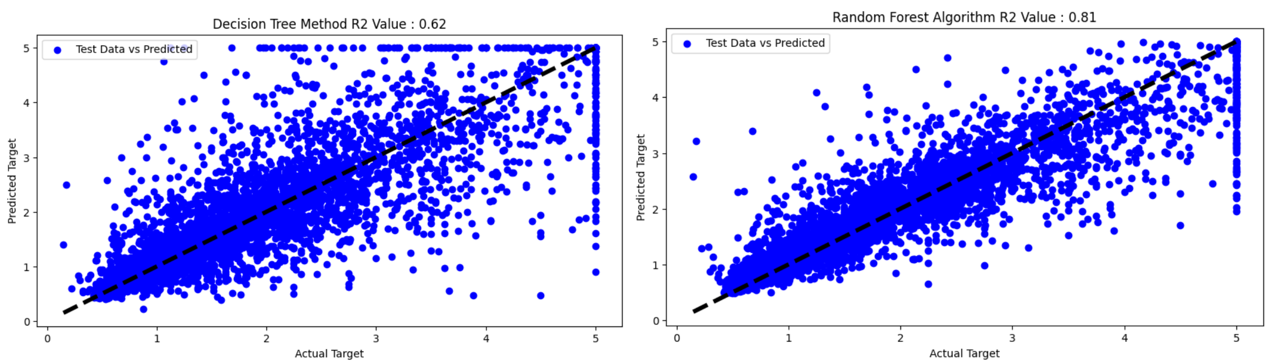

FIGURE 3. (a) Decision Tree Method R^2 Value Graph Random Forest Algorithm R^2 Value Graph

In Figure 3 (a), in the R^2 chart of house prices estimated using the Decision Tree method, the points are dispersed and diverge directly. Considering this graph and the R^2 value, it is concluded that the values estimated by the Decision Tree method are not very close to the actual values. Considering the R^2 graph of house prices estimated using the Random Forest Algorithm in Figure 3 (b), the convergence of the points to the line shows that the values estimated with the Random Forest Algorithm are closer to the actual values. As a result of the analysis of the obtained results and metrics, it was seen that the Random Forest Algorithm gave a better result and made a better prediction than the Decision Tree Algorithm for the dataset examined (please see, Figure 3).

5. CONCLUSIONS AND RECOMMENDATIONS

Housing has always been a fundamental requirement for people, and changes in the economy can affect the real estate market dramatically. Currently, housing values are on the rise, making accurate price estimation increasingly important. Inaccurate predictions can disrupt the market balance and lead to lost sales due to price differences in the free market. Fortunately, the development of technology has allowed for the effective use of artificial intelligence in various sectors, including finance, real estate, health, and production. AI solutions can prevent such problems and offer effective solutions.

In this study, machine learning, a sub-branch of artificial intelligence, was used and the Decision Tree method was used because it can process categorical-numeric data in the dataset, requires little data preparation, and is frequently used in multi-output problems. Random Forest Algorithm was used because it avoids overfitting problems, the Decision Tree method can produce trees that do not explain the data well, it identifies important features in the dataset used, and it is frequently used in regression-classification problems. The study examined and evaluated 20,640 data points, with each data point having eight features, from a dataset called "California Housing Prices".

The study analyzed various features and used performance evaluation metrics such as MSE, RMSE, R2, and MAE to compare the prediction accuracy of the Decision Tree and Random Forest algorithms. By comparing the metric values, the study determined which algorithm made better predictions on the dataset.

As an overall remark, it was observed that the Random Forest Algorithm outperformed the Decision Tree Algorithm in making predictions for the given dataset. The lower values of Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) and higher value of R-squared (R^2) indicate that the Random Forest Algorithm is more accurate in predicting the outcomes.

Price prediction was made by comparing the RMSE value of various algorithms using data from residences in Ames, Iowa [18]. The algorithm that made the best price prediction was determined by using different regression techniques such as Multilinear, Ridge, LASSO, Elastic Net, Gradient Boosting, Ada Boost Regression, and housing price prediction [19]. A housing rental price prediction model was developed with 5 different algorithms based on the rental housing prices in Yenimahalle district of Ankara and the characteristics of these houses [20]. It has been observed that there have been studies on house price forecasting in the literature, but there needs to be more studies in this field. Therefore, this study demonstrates the usefulness of machine learning algorithms as effective tools for producing accurate prediction results and establishes a solid basis for future projects. To further enhance the impact of this study, researchers can obtain data with more compelling features that are specific to a region at a national level, which will increase the sources of data for estimating housing prices. To address the issue of excessive pricing on housing throughout the countries, which is a significant problem, researchers can confidently develop projects that utilize different algorithms to reduce the prices of existing rents/sales without causing any harm to the homeowner in the long run.

Data Statement
The original data set used in the study can be accessed via the link;
https://scikit-learn.org/stable/datasets/real_world.html

Conflict of Interest
The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

REFERENCES

[1] Louati, A., Lahyani, R., Aldaej, A., Aldumaykhi, A., Otai, S., Price forecasting for real estate using machine learning: a case study on Riyadh city. *Concurrency and Computation: Practice and Experience*, *34*(6), e6748, (2022).

[2] Iban, M. C., An explainable model for the mass appraisal of residences: The application of tree-based Machine Learning algorithms and interpretation of value determinants. *Habitat International*, *128*, 102660, (2022).

[3] Yiğiter, Ş. Y., et al., Kira sertifikası fiyat değerlerinin makine öğrenmesi metodu ile tahmini. *International Journal of Islamic Economics and Finance Studies*, 4.3: 74-82 (2018).

[4] Acı, M.; Dogansoy, G.A, Demand forecasting for e-retail sector using machine learning and deep learning methods. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 37.3: 1325-1339 (2022).

[5] Barut, Z.; Bilgin, T. T., Konut Fiyatlarının Tahmini için Polinomsal Regresyon ve Yapay Sinir Ağları Yöntemlerinin Uygulamalı Karşılaştırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 27.1: 152-159 (2023).

[6] Burhan, H. A., Konut Fiyatları Tahmininde Makine Öğrenmesi Sınıflandırma Algoritmalarının Kullanılması: Kütahya Kent Merkezi Örneği. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi* 76: 221-237 (2023).

[7] Ozdemir, M.; Yıldız, K.; Büyüktanır, B., Housing price estimation with deep learning: a case study of Sakarya Turkey. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, 9.1: 138-151 (2022).

[8] İslamoğlu, E.; Bulut, H., Nevşehir ili konut fiyatlarını etkileyen faktörlerin hedonik fiyat modeli ile incelenmesi, *1st International Symposium of Silk Road Academic Studies*, (2018).

[9] Keleş, Ş.; Atabeyli, O.C., Konut ve hedonik fiyat bir yapısal eşitlik modeli uygulaması. *The International New Issues in SocialSciences*, 6.2: 111-128 (2018).

[10] Aydemir, E.; Aktürk, C,; Yalçınkaya, M.A., Yapay zekâ ile konut fiyatlarının tahmin edilmesi. *Turkish Studies*, 15.2: 183-194 (2020).

[11] Akay, E.Ç., et al., Türkiye konut fiyat endeksi öngörüsü: arıma, rassal orman ve arıma-rassal orman. *Press Academia Procedia*, 10.1: 7-11 (2019).

[12] Oral, M.; Okatan, E.; Kırbaş, İ., Makine öğrenme yöntemleri kullanarak konut fiyat tahmini üzerine bir çalışma: Madrid örneği. *3 rd International Young Researchers Student Congress*, p. 263-272 (2021).

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

[14] Gülağız, F.K.; Ekinci, E., Farklı Regresyon Analizi Yöntemleri Kullanılarak Ev Fiyatlarının Tahmini. *Conference: International Symposium on Industry*, p. 203-207 (2017).

[15] Özmaden, M.Ş; Erdal, M., Konut yapılarının maliyet tahmininde kullanılan yöntemlerin performans analizi. *Konya Journal of EngineeringSciences*, 8.4: 970-985 (2020).

[16] Gökalp, Ö. M. Makine öğrenmesi. *Gazi Üniversitesi, Gazi BilişimEnstitüsü, Adli Bilişim Bölümü* (2022).

[17] Pal, Nikhil R., et al., A possibilistic fuzzy c-means clusterin galgorithm. *IEEE transactions on fuzzy systems*, 13.4: 517-530 (2005).

[18] Fan, C.; Cui, Z.; Zhong, X., House prices prediction with machine learning algorithms. *Proceedings of the 2018 10th International Conference on Machine Learning and Computing*. p. 6-10(2018).

[19] Madhuri, CH Raga; Anuradha, G.; Pujitha, M. Vani., House price prediction using regression techniques: A comparative study. In: *2019 International conference on smart structures and systems (ICSSS)*. IEEE, p. 1-5(2019).

[20] Özşahin, M; Yılmaz, K.B; Akcan, S., Mühendislikte Araştırma ve Değerlendirmeler II: Makine öğrenmesi algoritmalari ile ev kirasi tahmini, *Gece Kitaplığı Yayınevi*, 2022