A Study on the Prediction of Apartment Prices using the GBRT model: A Case Study in Vinh City, Vietnam

Ha-Lan Tran
Nghe An University of Economics, Vietnam
tranhalan@naue.edu.vn

Thuy-Linh Tran Thi
Department of Civil Engineering, Vinh University, Vietnam
tranthuylinhlinh2004@gmail.com

Thanh-Vu Tran
Department of Civil Engineering, Vinh University, Vietnam
thahtrvu201@gmail.com

Doan-Huong Doan Thi
Department of Civil Engineering, Vinh University, Vietnam
doanhuongdhv@gmail.com

Trong-Ha Nguyen
Department of Civil Engineering, Vinh University, Vietnam
trongha@vinhuni.edu.vn

Received: 1 April 2024 | Revised: 19 April 2024 | Accepted: 22 April 2024

ABSTRACT

This study aims to propose an efficient Machine Learning (ML) model, namely Gradient Boosting Regression Trees (GBRT), to predict apartment prices considering the fluctuation of construction material prices and the annual inflation index. For developing the ML model, 480 apartments in Vinh City (Vietnam) were considered. The input parameters employed while training the ML model were the area of the apartments, the number of bedrooms/restrooms, the apartment class, nearby health or education services, investment potential, and parking, whereas the apartment price was the output of the model. The results show that the GBRT model predicts the apartment price accurately with a high value of 0.997 and a small RMSE of 0.26. Additionally, the obtained a20-index is very high, almost 1.0. Finally, a practical graphical user interface was developed to facilitate the prediction of the apartment price in terms of usability.

Keywords-GBRT model; apartment price; GUI; Vinh City; Vietnam

I. INTRODUCTION

The prediction of apartment price is not only a useful tool for managing risks and making investment decisions, but also helps to better understand the true value of real estate and market trends. Apartment prices are influenced by many factors such as the apartment area, project location, surrounding amenities, economic growth rate, and inflation. There are many factors affecting apartment sales and rental prices, namely the interest rates, inflation rates, economic growth rate, economic development level, industrialization rate, financial market developments, environmental factors, and building characteristics [1-5]. Recently, Artificial Intelligence (AI) algorithms have been widely applied in engineering techniques [6-12] and, in particular, the application of AI models in apartment price prediction has been the subject of many studies. In [13], Machine Learning (ML) was used to predict the real-time apartment prices in Tokyo metropolitan area. An ML approach to predict apartment prices was developed deploying a ML approach for Colombia in [14, 15] and Czech Republic in [16]. Even though various ML models were employed for estimating apartment prices, the Gradient Boosting Regression Trees (GBRT) model has not been utilized yet. Specifically, a consideration of various influential parameters including apartment area, existence of bedroom/restroom, apartment class, health services, education services, investment potential, and parking have not been investigated so far.

This study focuses on the application of the GBRT model to estimate apartment prices considering the fluctuation of construction material prices and the annual inflation index. To develop the proposed model, this study conducted an extensive survey of apartment prices in Vinh city, Vietnam, considering a total of 480 apartments in 11 projects. In particular, 7 input factors affecting apartment prices were considered and apartment price was the output variable of the GBRT model. Finally, an apartment price prediction model that takes into account fluctuations in material prices and inflation index was implemented. Additionally, a graphical user interface software was built in MATLAB. Furthermore, the quantitative influence of factors on the price of apartments was also contemplated.

II. MATERIALS AND METHODS

A. *Gradient Boosting Regression Trees (GBRT)*

The GBRT model is a powerful method in the field of ML used in classification and regression problems. GBRT combines the strength of decision trees and boosting techniques to create a flexible and accurate prediction model. The critical components of the algorithm are:

1) Boosting

GBRT belongs to the family of "boosting" algorithms, which is a strategy that focuses on building a sequence of weak models and combining them to create a stronger model. Each weak model in the chain learns from the errors of the previous model, thereby increasing the prediction accuracy.

2) Regression Trees

The sub-models in GBRT are decision trees. Decision trees are built by dividing data into groups based on characteristic variables, creating complex decision rules.

3) Gradient Descent

To optimize the model, GBRT engages the gradient descent method. Each tree is built to reduce the prediction error of the previous model, and gradient descent is applied to adjust the weights of the tree, reducing the magnitude of the gradient and making the model converge more quickly.

4) Loss Function

The loss function measures the difference between the predicted value and the actual value. For apartment price forecasting problems, the Mean Squared Error (MSE) value is often deployed to evaluate the error of predicted models.

5) Regularization

To avoid the overfitting phenomenon during the data training process, GBRT applies regularization algorithms. Parameters, such as number of trees, tree depth, and learning rate can be adjusted to control overfitting.

B. *The GBRT Model for predicting Apartment Prices*

The flowchart of the GBRT model is illustrated in Figure 1.

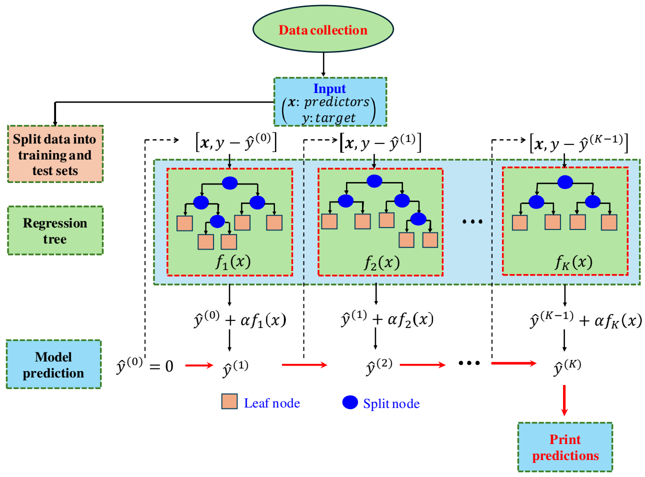

Fig. 1. The training process of the tree-based model.

The detailed steps for implementing the model are:

- Step 1. Initiation of GBRT

Create the initial prediction $\hat{y}_i$, which is the mean value of the output (i.e. apartment price).

- Step 2. Boosting loop (k from 1 to K)
  - Step 2.1: Calculate residuals. The ccalculated error is $r_{ki} = y_i - \hat{y}_{ki-1}$, where $\hat{y}_{ki-1}$ is the prediction of the previous tree.
  - Step 2.2: Develop the regression tree. Use training data and residuals to construct a new decision tree $\left(f_k(x)\right)$.
  - Step 2.3: Optimize the tree. Apply the optimization technique to reduce errors efficiently.
  - Step 2.4: Update prediction. Update the model predictions: $\hat{y}_{ki} = \hat{y}_{ki-1} + \alpha f_k(x)$, where α is the learning rate.
  - Step 2.5: Calculate error. The calculated error $L(y_i, \hat{y}_{ki})$ is obtained based on the new prediction.
- Step 3. Total prediction: $\hat{y}_i = \sum_{k=1}^{K} \hat{y}_{ki}$.
- Step 4. Performance evaluation and fine-tuning of hyper-parameters. Evaluate the model performance on the test set and possibly perform fine-tuning steps to improve predictions and reduce overfitting. Parameters, such as number of trees, tree depth, and learning rate can be adjusted to control overfitting.
- Step 5. Application. Appy the optimal GBRT for predicting new data.

C. *Infuential Parameters on Apartment Prices*

Determining the price to sell or rent an apartment is a complex process that depends on many factors. In this study, potential factors affecting the price of apartments are adapted from [4, 5, 17-20], including area X_1, number of bedrooms/restrooms X_2, apartment class X_3, health services X_4, education services X_5, investment potential X_6, and parking X_7 Those parameters are summarized in Table I.

TABLE I. FACTORS AFFECTING APARTMENT PRICES

Factors affecting	**References**
Area apartment X_1	[5, 17]
Number of bedrooms/restrooms X_2	[5, 17]
Apartment class X_3	[4]
Health services X_4	[4, 18, 19]
Education services X_5	[4, 18, 19]
Investment potential X_6	[5, 20]
Parking X_7	[4, 5]

D. Apply GBRT for predicting Apartment Prices

The application of the proposed GBRT model for predicting apartment prices is a 5-step process. Its block diagram can be seen in Figure 2. The steps are:

- Step 1. Select parameters affecting apartment prices
- Step 2. Collect and process data
- Step 3. Train the GBRT model
- Step 4. Evaluate the results
- Step 5. Apply the model.

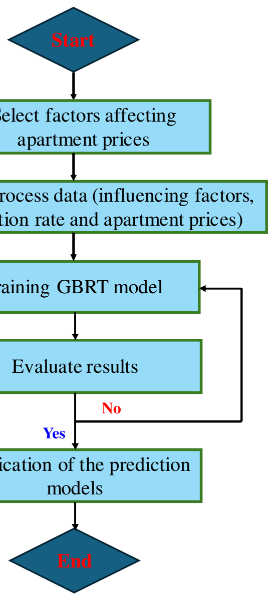

Fig. 2. Flowchart of the application of the proposed GBRT model.

III. CASE STUDY

A. Data Collection and Analysis

In this study, an investigation of the selling prices of 480 apartments in Vinh city, Vietnam in the period from 2018 to 2023 was carried out. The collected data include the apartment selling price and the considered factors $X_1 - X_7$ affecting that price. The price of the apartment is considered as the output. Specifically, apartment class X_3 is classified into three levels: level 1 (A class), level 2 (B class), and level 3 (C class). Health services (X_4), education services (X_5), and investment potential (X_6), are also divided into three levels: level 1 (excellent), level 2 (good), and level 3 (medium). Parking lot (X_5) is classified into three groups: sufficient parking space, not enough parking space, and very limited space. The statistical properties of the database are described in Table II, and the histograms of the dataset are portrayed in Figure 3. The correlation between the variables is displayed in Figure 4.

TABLE II. STATISTICAL SUMMARY OF THE DATASET

	Minimum	**Maximum**	**Mean**	**S.td**	**COV**
X_1(m^2)	30.6	112.0	64.6	15.6	0.2
X_2	1.0	3.0	2.3	0.6	0.3
X_3	1.0	3.0	2.3	0.7	0.3
X_4	1.0	3.0	1.7	0.7	0.4
X_5	1.0	3.0	1.8	0.6	0.3
X_6	1.0	3.0	1.9	0.8	0.4
X_7	1.0	2.0	1.4	0.4	0.3
Y(100$)	11.0	34.1	18.3	4.0	0.2

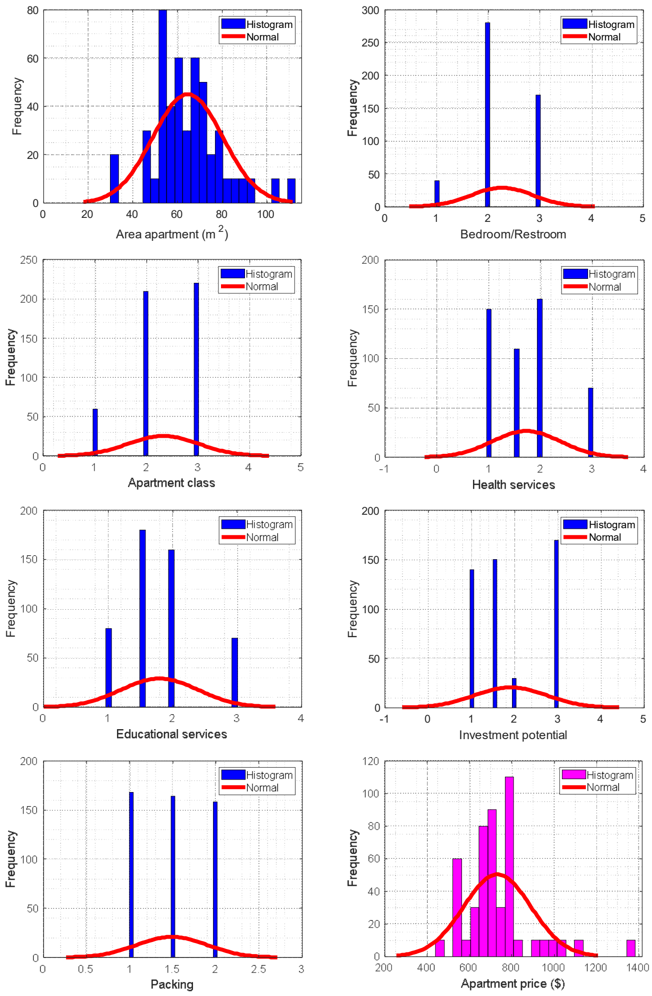

Fig. 3. Histograms of the input and output parameters.

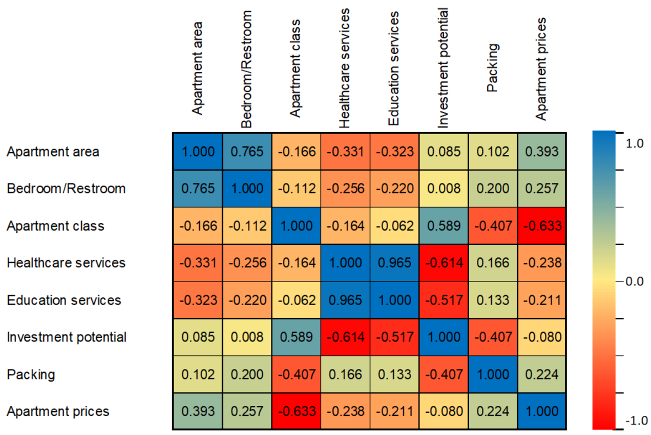

Fig. 4. Correlation between the variables.

B. Performance of the GBRT Model

Figure 5 manifests the regression of the developed GBRT model and the collected values, where the output predictions are very consistent with the collected data.

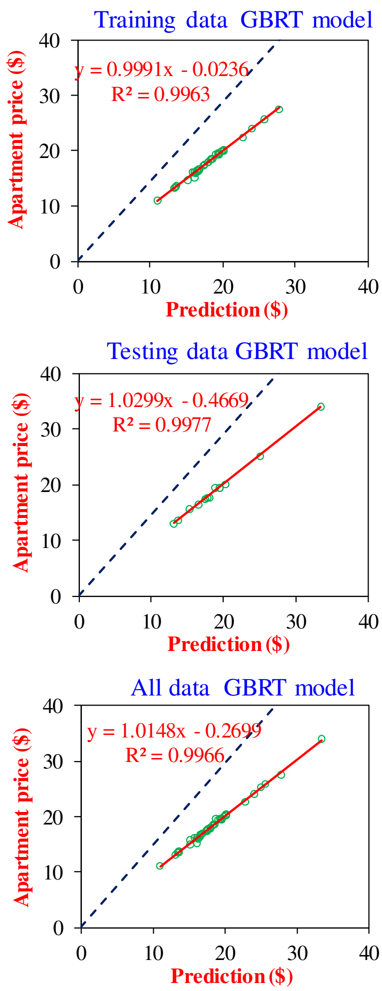

Fig. 5. Flowchart of the regression of predicted and collected data on apartment prices.

The R^2 values for training data, test data, and overall data are 0.9963, 0.9977, and 0.9966, respectively. It can be seen that those values are very close to 1.0. These results emphasize that the developed GBRT model can predict apartment prices accurately. The predicted result after training combined with the increment of market prices and annual inflation helps managers, investors, and, especially, clients accurately balance their prepared finances to buy an apartment. Figure 6 shows the convergence of the MSE and the comparison between the predicted and the collected apartment prices during data training. It can be noticed that the error values of the training and test data are close to zero after 1000 simulation trees. This result once again confirms that the GBRT model accurately predicts the apartment prices based on the collected data. Table IV provides the minimum, mean, maximum, Standard Deviation (SD) and Covariance of Variation (CoV) of the actual/predicted data ratio of the GBRT model. Mean values close to 1.0 and small SD values indicate that GBRT model has a superior ability in predicting apartment prices. A smaller CoV value implies a minimized amount of dispersion in the results.

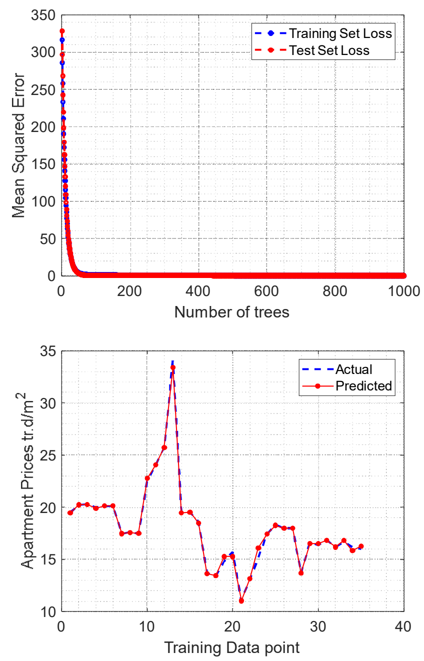

Fig. 6. MSE and comparison between predicted data and collected data.

TABLE III. CALCULATED RESULTS OF THE STATISTICAL PARAMETERS OF THE GBRT MODEL

	R^2	**a-20 index**	**RMSE**
Training	0.9963	1.00	0.2607
Testing	0.9977	1.00	0.3000
All data	0.9966	1.00	0.2401

TABLE IV. RESULTS OF THE RATIO OF ACTUAL (COLLECTED) TO PREDICTION PRICES

	Actual/prediction				
	Min	**Max**	**Mean**	**SD**	**CoV**
Training	0.9824	1.0613	1.0025	0.0128	0.0128
Testing	0.9659	1.0184	0.9960	0.0138	0.0139
All data	0.9659	1.0613	1.0007	0.0135	0.0134

C. The Influence of Input Factors on Apartment Prices

As analyzed above, the GBRT model has the ability to accurately predict apartment prices, which means that the contribution value of the input parameters reflects their influence on the output value. In this section, this study identifies the most influential factors that affect the price of the apartments. The influence of the input parameters on the output values is observed in Figure 7.

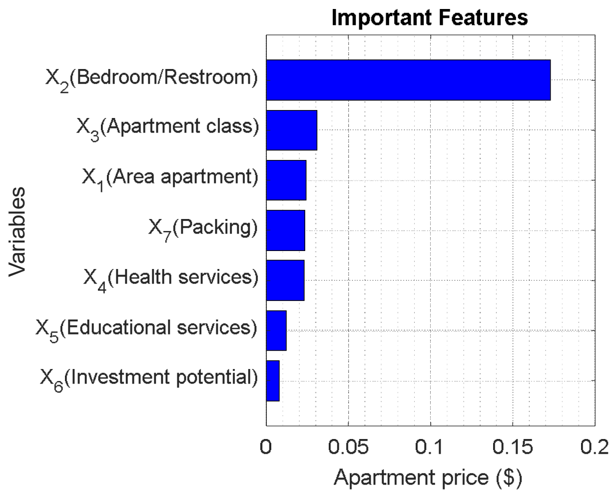

Fig. 7. Influence of the input variables on apartment prices.

D. Graphical User Interface

A convenient Graphical User Interface (GUI) tool was developed using MATLAB (Mathworks 2022b) to simplify the calculation of the apartment prices, as noticed in Figure 8.

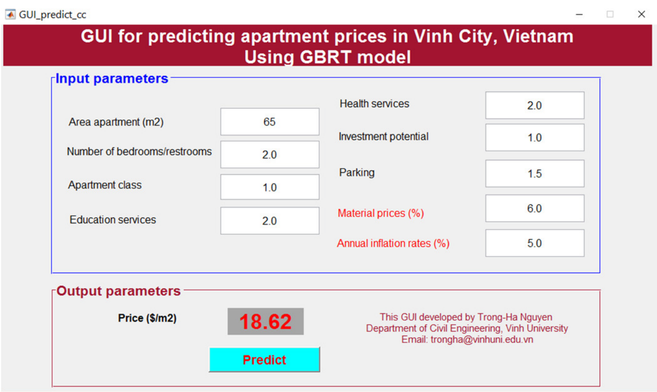

Fig. 8. GUI for apartment price.

In particular, the GUI has integrated additional variables predicting the fluctuation of material prices and the cumulative annual inflation index. Seven input parameters, from X_1 to X_7, and two price factors, i.e. volatility and inflation index, are included. Users can easily get the output by clicking on the "Predict" button. It should be noted that this GUI is developed based on the proposed GBRT model, so the verification of the prediction results has been thoroughly completed. However, this model cannot solve the extrapolation problem. So the input values must be limited to the minimum and maximum of the database. To expand the coverage of these ML models, it is necessary to consider a broader range of data used.

IV. CONCLUTIONS

This study developed a GBRT model to predict apartment prices considering construction material prices and annual inflation index. To develop the proposed model, this study conducted an extensive survey of the apartment prices in Vinh city, Vietnam with 480 apartments in 11 projects. In particular, seven factors affecting apartment prices were considered as input and the apartment price as the output of the GBRT model. The following conclusions were drawn:

- The GBRT model forecasts the apartment prices accurately with R^2 for training data and test data being 0.9963, 0.9977, respectively. The RMSE and a20-index of the training and test data sets are (0.2607, 1.0) and (0.300, 1.0), respectively.
- The influence of the considered factors on the predicted price was evaluated. The number of bedrooms/restrooms is the most influential parameter, followed by apartment class, whereas the development potential does not seem to affect the apartment price.
- A convenient graphical user interface was developed to simplify the calculation of apartment prices.

REFERENCES

[1] P. Abelson, R. Joyeux, G. Milunovich, and D. Chung, "Explaining House Prices in Australia: 1970–2003," *Economic Record*, vol. 81, no. s1, pp. S96–S103, 2005, https://doi.org/10.1111/j.1475-4932.2005.00243.x.

[2] H. K. Singla and P. Bendigiri, "Factors affecting rentals of residential apartments in Pune, India: an empirical investigation," *International Journal of Housing Markets and Analysis*, vol. 12, no. 6, pp. 1028–1054, Jan. 2019, https://doi.org/10.1108/IJHMA-12-2018-0097.

[3] M. Kamal and S. A. Pramanik, "Factors Affecting Customers to Buy Apartments in Dhaka City," *Daffodil International University Journal of Business and Economics,* vol. 9, no. 2, pp. 37–49, Dec. 2015.

[4] M. Hilmi, M. Masri, A. H. Nawawi, and I. Sipan, "Review of Building, Locational, Neighbourhood Qualities Affecting House Prices in Malaysia," *Procedia - Social and Behavioral Sciences*, vol. 234, pp. 452–460, Oct. 2016, https://doi.org/10.1016/j.sbspro.2016.10.263.

[5] H. Selim, "Determinants of house prices in Turkey: Hedonic regression versus artificial neural network," *Expert Systems with Applications*, vol. 36, no. 2, part 2, pp. 2843–2852, Mar. 2009, https://doi.org/10.1016/j.eswa.2008.01.044.

[6] D.-D. Nguyen and T.-H. Nguyen, "GBRT-based model for predicting the axial load capacity of the CFS-SOHS columns," *Asian Journal of Civil Engineering*, vol. 24, no. 8, pp. 3679–3688, Dec. 2023, https://doi.org/10.1007/s42107-023-00743-w.

[7] T.-H. Nguyen, N.-L. Tran, V.-T. Phan, and D.-D. Nguyen, "Prediction of shear capacity of RC beams strengthened with FRCM composite using hybrid ANN-PSO model," *Case Studies in Construction Materials*, vol. 18, Jul. 2023, Art. no. e02183, https://doi.org/10.1016/j.cscm.2023.e02183.

[8] T.-H. Nguyen, V.-T. Phan, and D.-D. Nguyen, "Practical ANN Model for Estimating Buckling Load Capacity of Corroded Web-Tapered Steel I-Section Columns," *International Journal of Steel Structures*, vol. 23, no. 6, pp. 1459–1475, Dec. 2023, https://doi.org/10.1007/s13296-023-00781-9.

[9] N.-L. Tran, D.-D. Nguyen, and T.-H. Nguyen, "Prediction of speed limit of cars moving on corroded steel girder bridges using artificial neural

networks," *Sādhanā*, vol. 47, no. 3, Jun. 2022, Art. no. 114, https://doi.org/10.1007/s12046-022-01899-y.

[10] F. Mlawa, E. Mkoba, and N. Mduma, "A Machine Learning Model for detecting Covid-19 Misinformation in Swahili Language," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10856–10860, Jun. 2023, https://doi.org/10.48084/etasr.5636.

[11] K. Theofilatos, S. Likothanassis, and A. Karathanasopoulos, "Modeling and Trading the EUR/USD Exchange Rate Using Machine Learning Techniques," *Engineering, Technology & Applied Science Research*, vol. 2, no. 5, pp. 269–272, Oct. 2012, https://doi.org/10.48084/etasr.200.

[12] R. F. Kamala and P. R. J. Thangaiah, "A Novel Two-Stage Selection of Feature Subsets in Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4169–4175, Jun. 2019, https://doi.org/10.48084/etasr.2735.

[13] T.-C. Peng and C.-C. Wang, "The Application of Machine Learning Approaches on Real-Time Apartment Prices in the Tokyo Metropolitan Area," *Social Science Japan Journal*, vol. 25, no. 1, pp. 3–28, Mar. 2022, https://doi.org/10.1093/ssjj/jyab029.

[14] T.-C. Peng and C.-C. Wang, "The Application of Machine Learning Approaches on Real-Time Apartment Prices in the Tokyo Metropolitan Area," *Social Science Japan Journal*, vol. 25, no. 1, pp. 3–28, Mar. 2022, https://doi.org/10.1093/ssjj/jyab029.

[15] A. A. Neloy, H. M. S. Haque, and Md. M. Ul Islam, "Ensemble Learning Based Rental Apartment Price Prediction Model by Categorical Features Factoring," in *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, New York, NY, USA, Oct. 2019, pp. 350–356, https://doi.org/10.1145/3318299.3318377.

[16] J. Nikodym, "Application of machine learning methods for estimating apartment prices in the Czech Republic," M.S. thesis, Univerzita Karlova, Prague, Czech Republic, 2019.

[17] G. Ekşioğlu Çetintahra and E. Çubukçu, "Çevre estetiğinin konut fiyatlarına etkisi," *İTÜDERGİSİ/a*, vol. 10, no. 1, Sep. 2011.

[18] S. O. Tze, "Factors affecting the price of real estate properties in Malaysia," *Journal of emerging issues in Economics, Finance and Banking*, vol. 1, no. 5, pp. 1–15, 2013.

[19] G. K. Babawale and Y. Adewunmi, "The Impact of Neighbourhood Churches on House Prices," *Journal of Sustainable Development*, vol. 4, no. 1, pp. 246–253, Jan. 2011, https://doi.org/10.5539/jsd.v4n1p246.

[20] F. Ersoz, T. Ersoz, and M. Soydan, "Research on Factors Affecting Real Estate Values by Data Mining," *Baltic Journal of Real Estate Economics and Construction Management*, vol. 6, no. 1, pp. 220–239, Dec. 2018, https://doi.org/10.2478/bjreecm-2018-0017.