Prediction of Apartment Price Considering Socio Economic and Crime Rates Factors in DKI Jakarta

David Noorcahya[1], Achmad Pratama Rifai[1], Agus Darmawan[1], Wangi P Sari[1]
[1]Department of Mechanical and Industrial Engineering, Universitas Gadjah Mada, Indonesia

Article Info

Article history:

Received May 26, 2023
Revised Jul 30, 2023
Accepted Sep 13, 2023

Keywords:

Machine Learning
Gradient Boosting
Apartment price
Socio Economic
Crime Rates

ABSTRACT

Investing in real estate properties in Indonesia is highly lucrative due to their consistent appreciation in value. Amongst the various property types, apartments are particularly favored for investment in limited land space. However, determining the value of apartments is often subjective and lacks quantitative measures. To address this issue, this study develops prediction models to predict rental prices and asset value based on apartment specifications, socio-economic factors, and crime rates. Machine learning models employed include Random Forest, Decision Tree, and Gradient Boosting Machine. The findings show Gradient Boosting Machine exhibits the highest accuracy in predicting apartment rental and sale prices, achieving R^2 values of 0.9230 and 0.8460, respectively. The study also highlights the significant influence of socio-economic factors and crime rates on the performance of the models, contributing between 0.09 and 0.22 with an average of 0.14, as indicated by the improved R^2 values. This study demonstrate that these models can be valuable tools for real estate investors and professionals seeking quantitative measures to determine the value of apartments. By incorporating apartment specifications, socio-economic factors, and crime rates, the models can provide objective insights into the potential rental income and asset value of apartments

Corresponding Author:

Achmad P. Rifai,
Department of Mechanical and Industrial Engineering,
Universitas Gadjah Mada,
Jl Grafika No, 2 Sleman DI Yogyakarta, Indonesia
Email: achmad.p.rifai@ugm.ac.id

1. INTRODUCTION

Indonesia, currently ranked as the world's fourth-most populous nation, experiences varying levels of population density across its regions, with DKI Jakarta, the capital city of Indonesia, having the highest density [1]. With its expanding population and rapid urbanization, the estate market in DKI Jakarta has continuously experienced significant growth and transformation [2]. As real estate prices continue to rise, property investment has become not only a necessity but also a promising avenue for economic growth. Among the various investment options, apartments are particularly attractive due to their vertical structure [3], maximizing housing availability in Jakarta's limited land space.

Socio-economic factors play a vital role in shaping the real estate market dynamics. DKI Jakarta, being the most developed city in Indonesia, attracts individuals seeking economic opportunities and improved living standards [4]. Factors such as income levels, employment rates, education levels, and amenities have a direct impact on the demand and pricing of apartments [5]. Moreover, the crime rate within a specific area can significantly influence the desirability and perceived safety of a location, affecting apartment prices [6].

Estimation of apartment values has been relied on subjective assessment and intuition due to the lack of quantifiable data. However, advancements in data collection and analysis techniques have provided unprecedented opportunities for researchers to leverage big data and machine learning algorithms to predict apartment prices accurately. By considering socio-economic and crime rate factors, researchers can gain insights into the relationship between these variables and apartment prices. This knowledge can assist investors in identifying areas with high growth potential and guide policymakers in formulating effective urban development strategies. Moreover, predicting apartment prices can help potential buyers and sellers make informed decisions and maximize their investment returns.

There are limited studies that have been done to predict apartment prices considering socio-economic and crime rate factors. Furthermore, to the best of our knowledge, there is none that predicts apartment prices in DKI Jakarta. Hence this study aims to fill the research gap and contribute to the understanding of dynamics that drive apartment prices in DKI Jakarta. In this study, we examine historical apartment price data, socio-economic indicators, and crime rate statistics in order to develop predictive models based on machine learning algorithms. These models can provide stakeholders in the real estate market with valuable insights and accurate forecasts, facilitating evidence-based decision-making and promoting sustainable growth in the apartment sector of DKI Jakarta.

2. RELATED WORK

The increasing demand for real estate and the volatile nature of the economy create a need to accurately predict real estate prices. Several studies in the literature have discussed the prediction of real estate prices using various predictive modelling techniques, such as linear regression [7]–[9], random forest [10]–[13], support vector machines [14]–[18], and metaheuristic algorithms [19]–[21].

Singh et al. [22] utilized dig data technologies to predict housing prices in Ames, Iowa, using sale data spanning from 2006 to 2010. Among the three models considered, gradient boosting demonstrates superior performance compared to linear regression and random forest. The latter method was also employed in another study to evaluate residentaial properties in Saint Petersburg, Russia [23]. The method demonstrated the ability to handle missing values and categorical variables effectively. Based on the empirical findings of this study, the random forest approach was found to perform better than other forecasting methods including multiple regression analysis and Artificial Neural Network. Another study by Neloy et al [24] investigated apartment price prediction based on ensemble learning. The study used a dataset encompassing factors such as construction year, price, area, type of apartment, amenities, and pet permits to predict apartment prices. Their findings revealed that the Random Forest model demonstrated the highest accuracy, achieving an RMSE value of 0.190286. Monika et al. [25] compared multiple machine learning models, including XGBoost, gradient boosting, random forest, light gradient boosting machine, and support vector regression. Their evaluation using MSE and RMSE indicated that the light gradient boosting machine model was the most effective in predicting housing prices.

Additionally, Tekin et al. [26] studied the effectiveness of various machine learning algorithms in reducing losses for investors. Their research highlighted the superiority of ensemble algorithms like random forest and XGboost in predicting property values compared to linear regression, polynomial regression, and decision trees. Moreover, Yağmur et al. [27] employed machine learning methods, such as artificial neural networks (ANN) and support vector regression (SVR), to predict house prices. Their study indicated that the ANN model outperformed the SVR and multiple linear regression models, providing accurate predictions that could benefit organizations involved in housing provision and valuation.

Several studies have highlighted the influence of socio-economic factors on real estate prices. Kang et al. [28] incorporated socio-economic characteristics such as population size, ethnicity, income, and unemployment rate to predict property values. Their study indicated that the Gradient Boosting Machine model successfully forecasted house price growth with an R^2 value of 0.74. Another study by Liu [29] investigated influencing factors of real estate prices based on linear regression model. The findings reveal strong strong positive correlations between the monthly income level of residents and the real estate price (r = 0.851) as well as per capita disposable income and the real estate price (r = 0.764). Additionally, a positive correlation is observed between the per capita housing expenditure of residents and the real estate price (r = 0.517), while a negative correlation is found between the completed area of real estate and the real estate price (r = 0.612).

The relationship between crime rates and real estate has been discussed in the literature. Previous research has examined the correlation between housing prices and the presence of crime hotspots identified through Getis-Ord statistics [30]. The study employed hedonic price modeling to evaluate the influence of crime hotspots on housing sales in the Stockholm metropolitan region, Sweden. While the impact of crime rates on house prices is generally modest, the finding of the study revealed a noticeable effect when considering proximity to crime hotspots. Moving a house just 1 kilometer farther away from a crime hot spot is shown to have a significant increase in its value. Similarly, another study examined the relationship between crime and housing prices also using hedonic regression [31]. The study considered varios vactors such as crime at the census tract level, changes in crime over time, differentiating between property crime and violent crime, and considering income levels of neighborhoods. The findings suggest that calculating average impact of crime rates on house prices can be misleader, as they are capitalized differently based per capita income (poor, middle class, and wealthy neigborhoods).

Despite the abundance of studies on predicting real estate values, there is a scarcity of research that incorporates socio-economic factors and crime rates. Moreover, there is a lack of research specifically focused on DKI Jakarta. This study seeks to address this research gap by investigating the effects of socio-economic factors and crime rates on the rental prices and selling prices of apartments in DKI Jakarta.

3. METHODS

The study comprises multiple stages: data collection, pre-processing, development of prediction models based on machine learning, and analysis. Initially, data are gathered for apartment specifications, apartment selling price index, socio-economic factors, and crime rates. Subsequently, the collected data undergoes pre-processing involving cleaning, handling missing values, and feature engineering. Exploratory data analysis is conducted to reveal insights and enhance understanding of the data. The subsequent state is employing the three machine learning models, i.e., random forest regression, decision tree regression, and gradient boosting machine. These models are evaluated using R^2, MAPE, and MSE. Finally, the results are analyzed, encompassing predictions for apartment rental prices, selling prices, and asset value growth.

3.1 Data Collection

The dataset, ranging from 2018 to 2022, was obtained from the websites Flokq (apartment specifications), Lamudi (historical data of apartment price), and Badan Pusat Statistik or Central Statistics Agency – (data related to socio-economic factors and crime rates). Apartment specifications data include rental price, square area, number of bedrooms and bathrooms, district, and city location. Whereas data related to socio-economic and crime rates factors include population ratio, workforce ratio, population density, proximity layout of apartment location to public facilities, urban murder and assault rate, and urban theft rate.

3.2 Pre-processing

In the pre-processing stage, as shown in Figure 1, the initial task is to clean the dataset by eliminating inconsistencies and errors. Subsequently, noise handling is carried out to identify and address invalid data. The next step involves handling missing values, where empty or missing data points are detected and dealt with accordingly. Categorical data handling is then employed to convert string and categorical data into a numeric format suitable for machine learning models.

Lastly, feature engineering is performed to extract additional information from the dataset with the aim of constructing effective machine learning models.

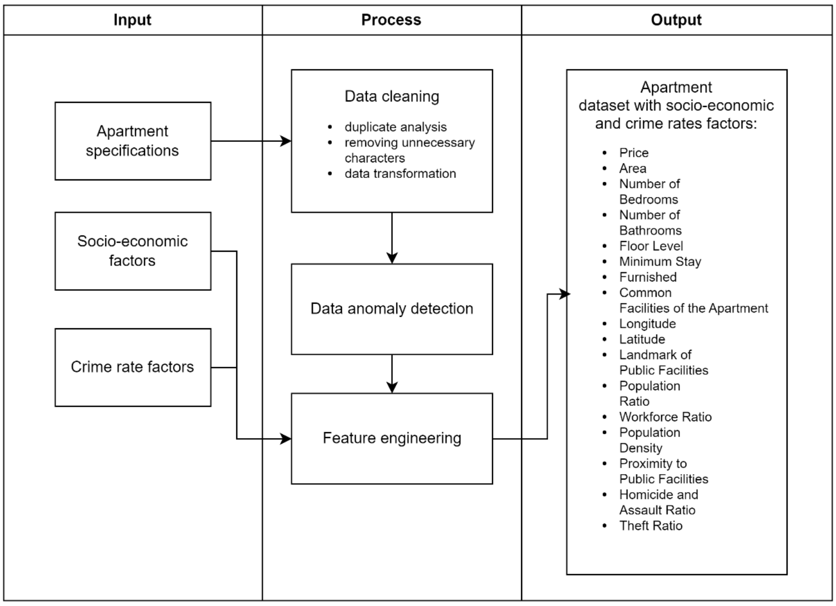

Figure 1. Pre-processing stages of machine learning methods employed to predict apartment prices in DKI Jakarta

3.3 Development of prediction model using machine learning

The next step is to employ machine learning modelling using Gradient Boosting Modelling (GBM), random forest (RF), and decision tree. This study employes predictive models based on regression and time series. The conceptual model for the regression and time series are depicted in Figure 2 and Figure 3 respectively.

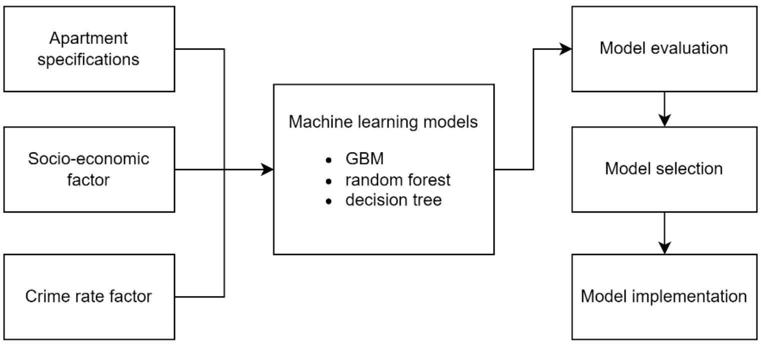

Figure 2. Conceptual model for regression prediction to predict apartment price in DKI Jakarta

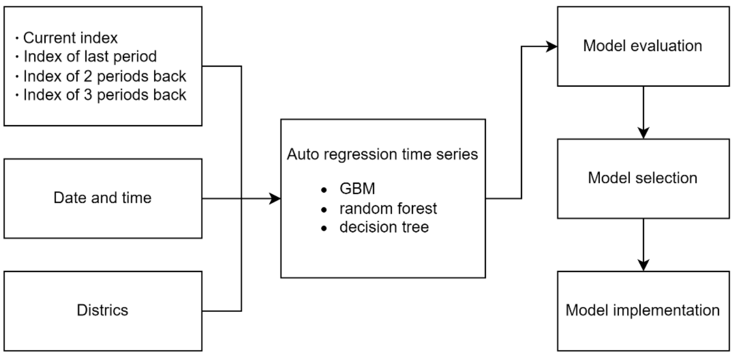

Figure 3. Conceptual model for time series prediction to predict apartment price in DKI Jakarta

The modeling stages for machine learning based on regression and time series are shown in Figure 4 and Figure 5 respectively. The figures explain the input, process, and output of each case.

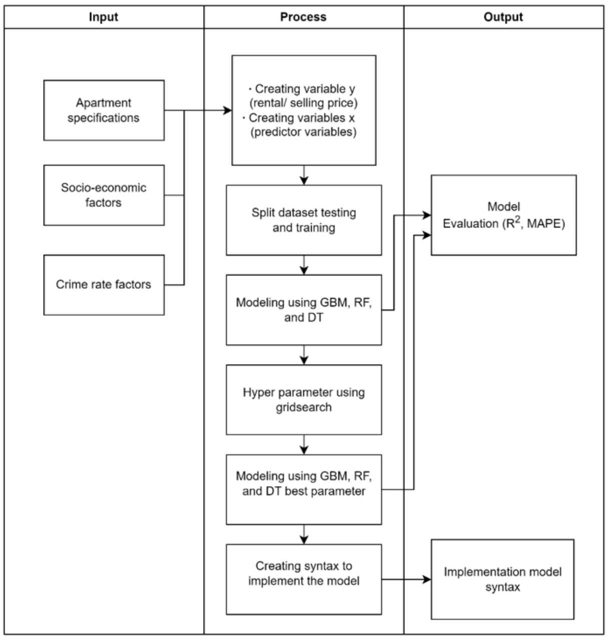

Figure 4. Input-Process-Output model of the machine learning regression to predict apartment prices in DKI Jakarta

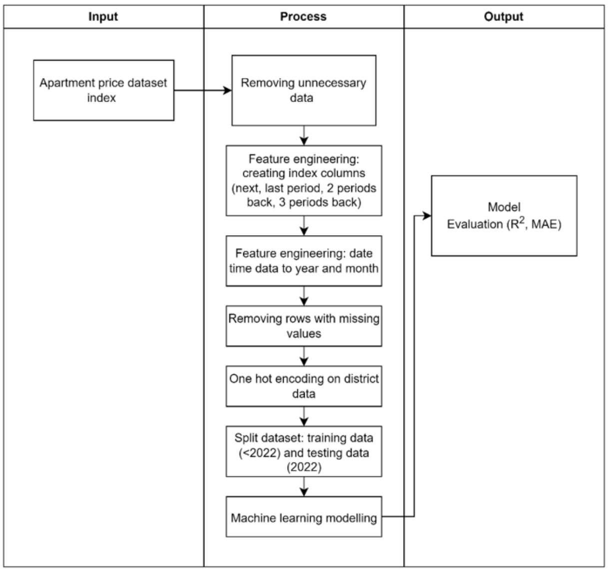

Figure 5. Input-Process-Output model of machine learning time series to validate prediction of apartment prices in DKI Jakarta

4. RESULTS AND DISCUSSION

Based on the collected data, the average rental price for an apartment is highest in South Jakarta, while the lowest average apartment prices are found in East Jakarta. Apartments located in South Jakarta generally have larger room sizes, more bedrooms, and more bathrooms compared to apartments in other areas. On the other hand, apartments located in East Jakarta tend to have smaller room sizes compared to other areas. Similarly, apartments located in South Jakarta have the highest average selling prices whilst apartments in East Jakarta have the lowest average selling prices. Notably, apartments in West Jakarta have the fewest average number of bedrooms and bathrooms among the areas mentioned. Whereas data for socio-economic and crime rates in DKI Jakarta can be seen in Figure 6.

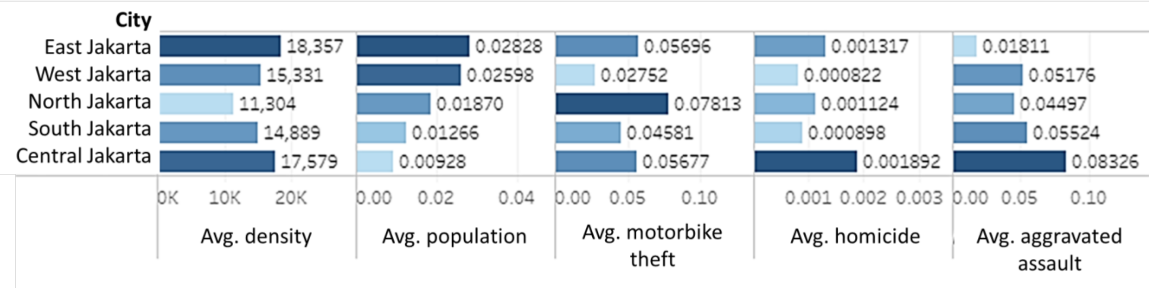

Figure 6. Socio-economic and crime rate in DKI Jakarta

4.1 Apartment Rental and Selling Prediction Modeling

Apartement rental prediction is initially conducted using default parameters. Subsequently, hyperparameter tuning process is implemented using GridSearchCV. The default and tuned parameters are depicted in Table 1.

Table 1. Default and tuned (after hyperparameter tuning) for renting apartments

Random Forest			**Decision Tree**			**Gradient Boosting Machine**		
Parameter	**Default**	**Tuned**	**Parameter**	**Default**	**Tuned**	**Parameter**	**Default**	**Tuned**
max_depth	None	80	max_depth	None	60	learning_rate	0.1	0.08
min_samples_leaf	1	1	max_features	None	None	max_depth	3	3
min_samples_split	2	3	max_leaf_nodes	None	None	n_estimators	100	5000
n_estimators	100	5000	min_samples_leaf	1	5			
			min_samples_split	2	2			

Machine learning modeling is conducted using the default and optimal parameters after hyperparameter tuning process. As shown in Table 2, Random Forest regression model outperforms the other two models for the initial modeling phase (using default parameters). Whereas for modeling using optimal parameters, Gradient Boosting Machine shows superior performance compared to random forest and decision tree models. Notably, all three models show better performance when using optimal parameters.

Table 2. The results of the apartment rental modeling using default and optimal parameters

Model	**Default parameters**		**Optimal parameters**	
	R^2	**MAPE**	**R^2**	**MAPE**
Random Forest	0.8796	0.1768	0.8831	0.1637
Decision Tree	0.7959	0.2214	0.8116	0.2086
Gradient Boosting Machine	0.8218	0.1989	0.9230	0.133

Table 3 illustrates the performance difference of apartment rent price prediction models in the training and testing processes. It can be seen from the table that the models perform better in the training process compared to the testing process for all three models used.

Table 3. Comparison of rental training set and testing set performance

Renting Apartment	**R^2 Training**	**R^2 Testing**
Random Forest	0.9764	0.8827
Decision Tree	0.9050	0.8130
Gradient Boosting Machine	0.9879	0.9232

A similar process is conducted to predict apartment selling prices. Initial modelling is executed using the default parameters. After which, hyperparameter tuning process is implemented using GridSearchCV. The results are shown in Table 4.

Table 4. Default and tuned (after hyperparameter tuning) for selling apartments.

Random Forest			Decision Tree			Gradient Boosting Machine		
Parameter	**Default**	**Tuned**	**Parameter**	**Default**	**Tuned**	**Parameter**	**Default**	**Tuned**
max_depth	None	80	max_depth	None	20	learning_rate	0.1	0.01
min_samples_leaf	1	2	max_features	None	None	max_depth	3	7
min_samples_split	2	3	max_leaf_nodes	None	None	n_estimators	100	1000
n_estimators	100	1000	min_samples_leaf	1	4			
			min_samples_split	2	5			

The performance of the three models employed are depicted in Table 5. Using default parameters, Random Forest outperforms the other two models. Whereas when optimal parameters are employed, Gradient Boosting Machined shows superior performance compared to Random Forest and Decision Tree models. Additionally, all the three models perform better using optimal parameters than default parameters. These results are consistent with the rental prediction modeling previously.

Table 5. The results of the apartment selling modeling using default and optimal parameters

Model	Default parameters		Optimal parameters	
	R^2	MAPE	R^2	MAPE
Random Forest	0.8349	0.2517	0.8484	0.2504
Decision Tree	0.7262	0.2878	0.7637	0.2754
Gradient Boosting Machine	0.8193	0.2501	0.8460	0.2328

Table 6 showcases the performance discrepancy between the apartment sale price prediction models during the training and testing processes. Notably, the models performed better in the training process compared to the testing process across all three models.

Table 6. Comparison of selling training set and testing set performance

Selling Apartment	R^2 Training	R^2 Testing
Random Forest	0.9479	0.8472
Decision Tree	0.9027	0.7632
Gradient Boosting Machine	0.9623	0.8448

4.2 Socio-economic factors and crime rate influence

The socio-economic factors used are population, workforce, and population density. Meanwhile, the crime rate factors used are the murder rate, rate of minor assault, rate of serious assault, motorbike theft, and general theft. Table 10 presents a comparison of the machine learning models' performance in predicting apartment rental and selling prices, considering the inclusion and exclusion of socio-economic and crime rate features. The result demonstrates that the models incorporating socio-economic and crime rate factors exhibit superior performance compared to the models without these features. The results for both renting and selling prices, which indicate the significant impact of socio-economic and crime rate factors on the predictive accuracy of the models. The findings highlight the significant impact of socio-economic and crime rate factors on the performance of the prediction models. By including these factors, the models are able to capture the influence of local economic conditions, demographic characteristics, and crime rates on apartment prices. This information is crucial for making accurate predictions and informed investment decisions in the real estate market.

Tabel 10. Performance comparison with and without socio-economic factors and crime rates on selling and renting prices

Machine learning models	Renting Prices		Selling Prices	
	Without	With	Without	With
Random Forest	0.7956	0.8827	0.7036	0.8472
Decision Tree	0.7070	0.8130	0.6326	0.7632
Gradient Boosting	0.7951	0.9232	0.6276	0.8448

The study also includes a feature analysis using SHAP (SHapley Additive exPlanations) values analysis for apartment rental and selling prices. This analysis provides insights into how each feature affects the prediction of apartment prices and the magnitude of their impact on apartment prices. Figure 7 depicts the SHAP values graph, which illustrates the influence of each feature on apartment prices. The analysis reveals that the size of the room has the greatest influence on the apartment price. This means that larger apartments tend to have higher prices. The number of bedrooms is the second most impactful feature, indicating that apartments with more bedrooms are generally priced higher. The location of the apartment, represented by longitude and latitude, also has a significant impact on the apartment price. Apartments located in the central and southern parts of Jakarta tend to be more expensive in terms of longitude, while apartments in the western part of Jakarta tend to be more expensive in terms of latitude. This suggests that the geographical positioning within Jakarta plays a role in determining apartment prices.

Regarding population and density features, the SHAP Values graph shows interesting findings. Higher population in an area generally corresponds to higher apartment prices, indicating that areas with a larger population tend to have a higher purchasing power and economic activity. However, the graph shows that high population density actually decreases apartment prices. This may be because high population density is often associated with slum areas or lower economic conditions. Additionally, the analysis highlights a correlation between the number of murders and motorcycle thefts in an area and higher apartment prices. This finding suggests that areas with higher crime rates may paradoxically have higher apartment prices, possibly due to factors such as increased security measures or demand from individuals who prioritize other aspects over safety.

Overall, the feature analysis using SHAP values provides valuable insights into the influence of various features on apartment prices. It reinforces the significance of apartment size, number of bedrooms, location, population, density, and crime rates as factors that impact apartment prices in Jakarta. Understanding these relationships can aid in decision-making processes for real estate investments and provide a more nuanced understanding of the factors affecting apartment prices in the region.

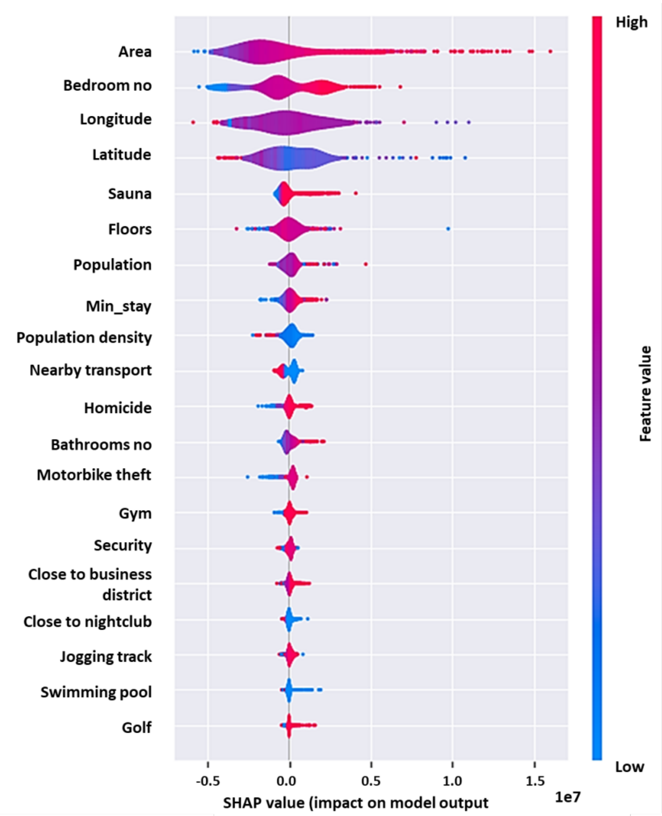

Figure 6. SHAP Values Analysis

5. CONCLUSION

The machine learning models used in the study show a high level of accuracy in predicting apartment prices. Specifically, the Gradient Boosting Machine reveals to be the best model for predicting both rental and sale prices, achieving R^2 values of 0.9230 and 0.8460, respectively. The study finds that the performance of the models was significantly influenced by socio-economic factors and crime rates, as indicated by their impact on the R^2 values. On average, these factors had a contribution of 0.14, with a range of 0.09 to 0.22. Most of the crime rate data in this study still uses city-wide data. In future research, you can add data on crime rates with a smaller scope, such as at the sub-district and village levels. Then, the rising and falling price trend data used in this study is data on the sub-district scope. In future research, you can add and use data from each apartment to make it more accurate.

The high level of accuracy achieved by the machine learning models in predicting apartment prices implies that these models can be valuable tools for real estate professionals, investors, and individuals looking to buy or rent apartments. The accuracy of the models, particularly the Gradient Boosting Machine, suggests that they can provide reliable estimates of apartment prices, both for rental and sale purposes. Since the study investigates the case at the city-wide leve, this may limit the granularity of the analysis and overlooks potential variations in crime rates within smaller geographical areas, such as sub-districts or villages. Future research can incorporate crime rate data at a smaller scale, such as sub-district and village levels. By including more localized crime data, the models can potentially capture finer variations in crime rates and their influence on apartment prices. This would provide a more comprehensive understanding of the relationship between crime rates and apartment prices, allowing for more accurate predictions.

REFERENCES

[1] BPS, “Kepadatan Penduduk menurut Provinsi (jiwa/km2), 2019-2021,” 2021. https://www.bps.go.id/indicator/12/141/1/kepadatan-penduduk-menurut-provinsi.html (accessed Jun. 09, 2023).

[2] E. Rustiadi, A. E. Pravitasari, Y. Setiawan, S. P. Mulya, D. O. Pribadi, and N. Tsutsumida, “Impact of continuous Jakarta megacity urban expansion on the formation of the Jakarta-Bandung conurbation over the rice farm regions,” *Cities*, vol. 111, p. 103000, 2021.

[3] M. Nethercote and R. Horne, “Ordinary vertical urbanisms: City apartments and the everyday geographies of high-rise families,” *Environ Plan A*, vol. 48, no. 8, pp. 1581–1598, 2016.

[4] R. A. Komalawati and J. Lim, “Reality of compact development in a developing country: focusing on perceived quality of life in Jakarta, Indonesia,” *International Journal of Urban Sciences*, vol. 25, no. 4, pp. 542–573, 2021.

[5] R. A. Rahadi, S. K. Wiryono, D. P. Koesrindartoto, and I. B. Syamwil, “Factors influencing the price of housing in Indonesia,” *International Journal of Housing Markets and Analysis*, 2015.

[6] V. Ceccato and M. Wilhelmsson, “The impact of crime on apartment prices: Evidence from Stockholm, Sweden,” *Geogr Ann Ser B*, vol. 93, no. 1, pp. 81–103, 2011.

[7] C. Ozgur, Z. Hughes, G. Rogers, and S. Parveen, “Multiple linear regression applications in real estate pricing,” *International Journal of Mathematics and Statistics Invention (IJMSI)*, vol. 4, no. 8, 2016.

[8] N. N. Ghosalkar and S. N. Dhage, “Real estate value prediction using linear regression,” in *2018 fourth international conference on computing communication control and automation (ICCUBEA)*, IEEE, 2018, pp. 1–5.

[9] A. Abdulhafedh, “Incorporating multiple linear regression in predicting the house prices using a big real estate dataset with 80 independent variables,” *Open Access Library Journal*, vol. 9, no. 1, pp. 1–21, 2022.

[10] S. Levantesi and G. Piscopo, “The importance of economic variables on London real estate market: A random forest approach,” *Risks*, vol. 8, no. 4, p. 112, 2020.

[11] R. Sawant, Y. Jangid, T. Tiwari, S. Jain, and A. Gupta, “Comprehensive analysis of housing price prediction in pune using multi-featured random forest approach,” in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, IEEE, 2018, pp. 1–5.

[12] J. Hong, H. Choi, and W. Kim, “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*, vol. 24, no. 3, pp. 140–152, 2020.

[13] S. Levantesi and G. Piscopo, “The importance of economic variables on London real estate market: A random forest approach,” *Risks*, vol. 8, no. 4, p. 112, 2020.

[14] K. C. Lam, C. Y. Yu, and C. K. Lam, “Support vector machine and entropy based decision support system for property valuation,” *Journal of property research*, vol. 26, no. 3, pp. 213–233, 2009.

[15] X.-J. Wang, G.-T. Zeng, K.-X. Zhang, H.-B. Chu, and Z.-S. Chen, “Urban Real Estate Market Early Warning Based on Support Vector Machine: A Case Study of Beijing.,” *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, pp. 153–166, 2020.

[16] W. Li, Y. Zhao, W. Meng, and S. Xu, “Study on the risk prediction of real estate investment whole process based on support vector machine,” in *2009 Second International Workshop on Knowledge Discovery and Data Mining*, IEEE, 2009, pp. 167–170.

[17] W. K. O. Ho, B.-S. Tang, and S. W. Wong, “Predicting property prices with machine learning algorithms,” *Journal of Property Research*, vol. 38, no. 1, pp. 48–70, 2021.

[18] R. Yang *et al.*, “Big data analytics for financial Market volatility forecast based on support vector machine,” *Int J Inf Manage*, vol. 50, pp. 452–462, 2020.

[19] M. Renigier-Biłozor, A. Chmielewska, M. Walacik, A. Janowski, and N. Lepkova, “Genetic algorithm application for real estate market analysis in the uncertainty conditions,” *Journal of Housing and the Built Environment*, vol. 36, no. 4, pp. 1629–1670, 2021.

[20] V. Del Giudice, P. De Paola, and F. Forte, “Using genetic algorithms for real estate appraisals,” *Buildings*, vol. 7, no. 2, p. 31, 2017.

[21] S. M. Muneer, M. B. Alvi, and M. A. Rasool, “Genetic Algorithm Based Intelligent System for Estate Value Estimation,” *International Journal of Computational and Innovative Sciences*, vol. 1, no. 1, 2022.

[22] A. Singh, A. Sharma, and G. Dubey, “Big data analytics predicting real estate prices,” *International Journal of System Assurance Engineering and Management*, vol. 11, pp. 208–219, 2020.

[23] E. A. Antipov and E. B. Pokryshevskaya, “Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics,” *Expert Syst Appl*, vol. 39, no. 2, pp. 1772–1778, 2012.

[24] A. A. Neloy, H. M. S. Haque, and M. 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*, 2019, pp. 350–356.

[25] R. Monika, “House Price Forecasting Using Machine Learning Methods,” *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 11, pp. 3624–3632, 2021.

[26] M. Tekin and I. U. Sari, “Real Estate Market Price Prediction Model of Istanbul,” *Real Estate Management and Valuation*, vol. 30, no. 4, pp. 1–16, 2022.

[27] A. Yağmur, M. Kayakuş, and M. Terzioğlu, “House price prediction modeling using machine learning techniques: a comparative study,” *Aestimum*, vol. 81, 2022.

[28] Y. Kang *et al.*, “Understanding house price appreciation using multi-source big geo-data and machine learning,” *Land use policy*, vol. 111, p. 104919, 2021.

[29] G. Liu, “Research on prediction and analysis of real estate market based on the multiple linear regression model,” *Sci Program*, vol. 2022, pp. 1–8, 2022.

[30] V. Ceccato and M. Wilhelmsson, “Do crime hot spots affect housing prices?,” *Nordic journal of criminology*, vol. 21, no. 1, pp. 84–102, 2020.

[31] G. E. Tita, T. L. Petras, and R. T. Greenbaum, “Crime and residential choice: a neighborhood level analysis of the impact of crime on housing prices,” *J Quant Criminol*, vol. 22, pp. 299–317, 2006.