BAREKENG: Journal of Mathematics and Its Applications
March 2025 Volume 19 Issue 1 Page 0107-0118
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss1pp0107-0118

MODELING HOUSE SELLING PRICES IN JAKARTA AND SOUTH TANGERANG USING MACHINE LEARNING PREDICTION ANALYSIS

Sugha Faiz Al Maula [1], Nicoletta Almira Dyah Setiawan[2], Elly Pusporani[3*], Sa'idah Zahrotul Jannah[4]

[1,2,3,4] *Statistics Study Program, Faculty of Science and Technology, Universitas Airlangga*
Jl. Dr. Ir. H. Soekarno, Surabaya, 60115, Indonesia

*Corresponding author's e-mail: * **elly.pusporani@fst.unair.ac.id***

Article History:
Received: 5th April 2024
Revised: 2nd September 2024
Accepted: 2nd September 2024
Published: 13th January 2025

Keywords:
Affordable housing;
Jabodetabek;
Machine learning regression.

ABSTRACT

The increasing demand for housing in urban agglomerations, particularly in areas like Jakarta, has made homeownership a significant challenge for many, especially first-time buyers and the lower-middle class. Post-pandemic shifts have further influenced housing preferences, driving interest towards suburban areas with green spaces. Despite government efforts through mortgage subsidy programs, affordability remains a concern, particularly in peripheral regions. This study aims to analyze housing prices in various Jakarta regions using machine learning models, including Multiple Linear Regression (MLR), Support Vector Regression (SVR), Light Gradient Boosting Machine (LGBM), and Random Forest. A dataset of 554 house prices from West Jakarta, South Jakarta, Central Jakarta, and South Tangerang was used. The analysis focused on key predictors like land area, building area, bedrooms, and carports, with R^2 and Mean Squared Error (MSE) metrics evaluating model performance. Results showed that LGBM and Random Forest outperformed others with 0.8 R^2 and low MSE, with building and land area as the most significant factors influencing prices. The study concludes that property size is a primary determinant of house prices, and there is a need for policy interventions to make housing more affordable. Additionally, apartment rentals offer a viable alternative, especially in central urban areas, where proximity to economic activities and facilities is crucial. The findings suggest that enhancing marketplace features with predictive tools could further assist buyers in making informed decisions.

How to cite this article:

S. F. A. Maula, N. A. D. Setiawan, E. Pusporani., and S. Z. Jannah, "MODELING HOUSE SELLING PRICES IN JAKARTA AND SOUTH TANGERANG USING MACHINE LEARNING PREDICTION ANALYSIS," *BAREKENG: J. Math. & App.*, vol. 19, iss. 1, pp. 0107-0118, March, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article • Open Access

1. INTRODUCTION

The goal of buying a house has become the dream of many people. Houses are purchased not only as a place to live but also as long-term investment assets for most people. For this reason, the increasingly narrow land area in large urban agglomeration areas, such as Jakarta, means that many workers from suburban areas are unable to buy houses in the city center. This condition became different after the pandemic, and housing preferences changed towards areas that were friendly to social infrastructure, suburban areas, and close to green open spaces [1].

In addition, houses are a primary need, and the massive urbanization occurring in big cities such as Jakarta has resulted in the lower middle class experiencing a backlog situation or a gap between the houses being built and what the community needs. The government has begun to address this problem through a program under the authority of the Ministry of Public Works and Housing; however, the housing mortgage subsidy program has not been sufficiently resolved [2]. The government has not been able to introduce this program and does not prioritize the lower-middle class, where it is very important to own a house, such as young families or newly married people [3]. Because, the program targeting peripheral area and the metropolitan core experiencing low population growth due to high cost of living then the peripheral areas are growing in both population and economic growth [4].

The conditions described above make house prices increasingly high. The factors for purchasing a house with first-time buyers include location, economic conditions, design, and facilities provided [5]. In addition to buying a house, people have alternatives, including renting an apartment. Research shows that there is no reciprocal causality between the selling price of an apartment and the rental price, meaning that changes in price in one factor do not necessarily result in changes in the other factors [6]. The rise of house price effect positive to consumption but in little effect for homeowners. Meanwhile the dropping of house price tend to negative effect to homeowners and positive effect to renters according to [7]. The factors for purchasing an apartment are different in each big city; for example, investment tends to be the main factor in Bandung. In Jakarta and Surabaya, location is the main factor [8]. Other research shows that people who live in apartments are 3.8 times more likely to be close to facilities such as education than those who live in landed houses [9]. This indicates buying/renting an apartment, because its location in the middle of the city is something that is taken into consideration when deciding where to live.

In recent years, due to the increasing trend towards big data, machine learning has become an important prediction approach as it can more accurately predict house prices based on their attributes, regardless of data from previous years. Several studies have explored this issue and proved the capabilities of machine learning approaches. Apart from the development of big data, another element of buying and selling, namely the marketplace, is not immune to developments in technology. With current developments, the marketplace has made information about buying and selling houses more open, easy, and free to obtain. As a result, prospective buyers have many choices in purchasing and renting all types of residences. Many choices can be refined if the marketplace provides additional features as suggestions or recommendations for potential buyers. The current marketplace function is used only as a search engine to search for houses for sale. The purpose of this article is to provide suggestions for marketplaces for additional features, determine the most accurate comparison of predictive models from the data obtained using various regression analyses in machine learning, and provide recommendations for home buyers or sellers in an effort to achieve the best model in the case of house prices [10].

2. RESEARCH METHODS

In this section, the experimental approach used for the regression is presented. Figure 1 shows the research flow, which consists of four methods. The first research flow was related to the data input and cleaning. The researcher then performed preprocessing and explained the regression process. The algorithm and data output are also explained at the end of the research flow.

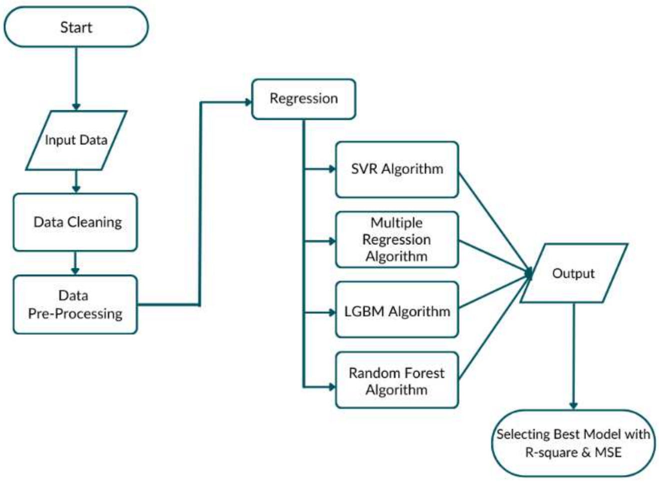

Figure 1. Research Flowchart

The steps for conducting the analysis in this study are as follows:

1. Carry out descriptive analysis of the data
2. Splitting data

The splitting data can be seen in the following table with a total of 554 data houses used.

Table 1. Splitting Data

No.	Region	Training	Testing	Predicting	Total
1	South Tangerang	122	31	10	153
2	South Jakarta	110	28	10	138
3	West Jakarta	107	26	10	133
4	Central Jakarta	104	26	10	130
	Total	443	111	40	554

Based on Table 1, it is known that the distribution of house data collection is quite even, around 130-150 houses in each region, and uses k-fold with k=5.

1. Formation of a regression model using the three methods mentioned above.
2. Performance or measure of the goodness of the model

   The model's goodness can be measured by a R^2 score that describes the extent to which the predictor variable can explain the response variable and the Mean Squared Error (MSE) which provides an indication of how far off the model's predictions are from the actual values, with larger errors being penalized more.

3. Interpret the ordinal logistic regression model obtained

   The regression model was interpreted by examining the R^2 score. The higher the score, the better is the score and for lower MSE indicates better model performance The scores were compared for all models used.

2.1 Dataset and Pre-Processing

The data used in this research is secondary data obtained by conducting a manual search via the website www.rumah123.com. The objective of this research is 554 data on house prices in West Jakarta, South Jakarta, Central Jakarta, and South Tangerang in 2024. The dataset used in this research contained attributes such as Bedroom (numerical), Bathroom (numerical), Carport (numerical), Land Area (numerical), Building Area (numerical), and One Hot Encoding for every region area (numerical). Before using the dataset, data cleaning was performed by removing outliers, standardization data, and filtering based on price location (West Jakarta, South Jakarta, Central Jakarta, and South Tangerang) and it is necessary to perform data preprocessing to reduce the noisy of the data without affecting the regression task [11]. Data pre-processing in this research used one hot encoding for region area with the aim of representing qualitative predictors for every region.

2.2 Methodology

After data cleaning and pre-processing, the next stage is to divide the data into k-fold data. Then, the house price dataset is regressed using SVR, Multiple Linear Regression, LGBM, and Random Forest methods. Finally, the model will predicting data with exclude from model.

The machine learning method used in this study was regression by predicting house prices using the predictor variables mentioned above. In this study, the analysis was performed using Multiple Linear Regression methods, Support Vector Regression (SVR), Light Gradient Boosting Machine (LGBM), and Random Forest. Multiple Linear Regression is a statistical technique to predict the result of an answer variable, using a number of explanatory variables. The object of (MLR) is to model the linear relationship between the independent variables x and dependent variable y that will be analyzed [12]. The basic model for MLR is:

$$y = B_0 + B_1X_1 + B_2X_2 + \cdots B_mX_m + \varepsilon$$

Support Vector Regression uses linear kernel functions for regression which is similar to support vector machines but SVR sets the tolerance margin (ε) to approximation not like SVM which should be taken from the problem is represented in Figure 2.

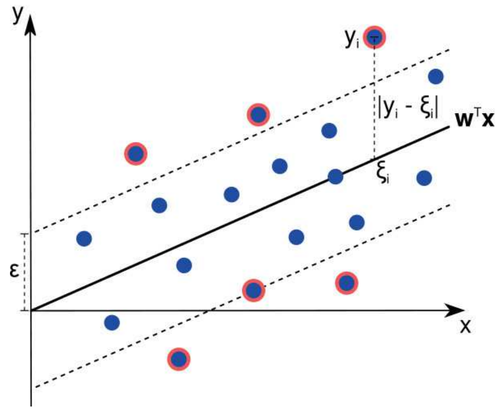

Figure 2. Support Vector Linear Regression

While SVM is used for classification tasks, where the goal is to find a hyperplane that separates classes with a maximum margin, SVR modifies this approach for regression by focusing on fitting a curve within a tolerance margin around the data points [13].

The LGBM algorithm enhances gradient-boosted decision tree models by improving runtime efficiency and reducing memory usage while preserving strong accuracy. Unlike other tree algorithms that expand horizontally, LGBM grows vertically through a leaf-wise approach, which helps in minimizing overfitting. While traditional algorithms expand trees level by level, LGBM focuses on the leaf with the greatest potential for reducing loss. By expanding the same leaf, LGBM achieves a more significant reduction in loss compared to the level-wise method. This approach of growing trees leaf-by-leaf, targeting the leaf with the highest delta loss, allows LGBM to produce better results with fewer trees, as it grows deeper trees that effectively minimize loss [14].

Figure 3. LGBM Illustration

Random Forest is an ensemble learning technique that constructs numerous decision trees, with each tree serving as an independent regression model. The final prediction in RF regression is obtained by averaging the outputs of all these decision trees, which helps enhance accuracy and mitigate overfitting. The core component of a Random Forest is the decision tree, also referred to as a Classification and Regression Tree (CART). In regression tasks, a decision tree iteratively divides the data into subsets based on specific criteria, refining predictions at each step. By aggregating the outputs from multiple decision trees, Random Forest produces more accurate and stable predictions, making it highly effective for various regression tasks, especially when handling complex or noisy datasets [15]. Besides regression methods, variable importance is a crucial metric in machine learning, especially in ensemble methods like random forests. It quantifies the contribution of each feature to the model's predictive power, typically ranging from 0 to 1. This metric helps identify which variables most significantly influence the model's predictions, allowing for better model interpretation.

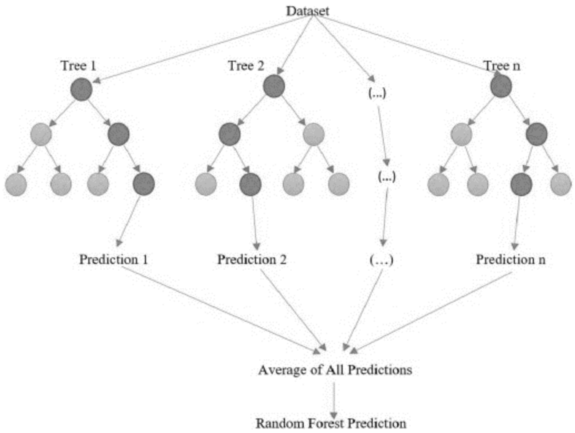

Figure 4. Random Forest Illustration

3. RESULTS AND DISCUSSION

In the results and discussion stage, the house price dataset was divided with 5-fold. This study aims to determine the results of regression analysis using SVR, Regression, LGBM, and Random Forest algorithms on the house price dataset. The four algorithms were compared based on accuracy measures to obtain the best algorithm based on R^2 and MSE .

3.1 Descriptive Statistics

When conducting data analysis, descriptive analysis can be used to describe the characteristics of the data used. A general description of the data is as follows.

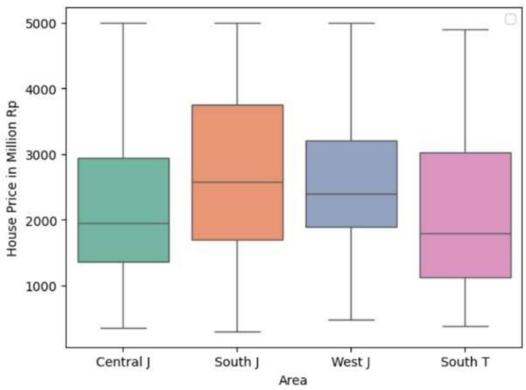

Figure 5. Area Boxplot

The boxplot in **Figure 5** illustrates the distribution of housing prices across Central Jakarta, South Jakarta, West Jakarta, and South Tangerang. South Jakarta stands out as the most expensive region, with average price around 2,700 million rupiah and a wide range of prices extending from about 1,500 million to nearly 5,000 million rupiah, indicating a market with numerous high-end properties. South Tangerang, the most affordable region, has average price around 1,900 million rupiah, with prices ranging between 1,000 million and 3,000 million rupiah, appealing to buyers looking for budget-friendly options. Overall, the boxplot highlights clear distinctions in housing price distributions, with South Jakarta catering to a higher-end market, while South Tangerang offers more accessible prices, reflecting the socio-economic dynamics and desirability of each location.

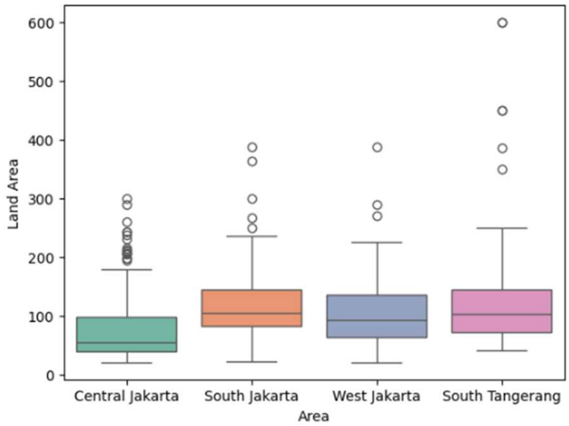

Figure 6. Land Area Boxplot

Figure 6 presents the distribution of land areas across Central Jakarta, South Jakarta, West Jakarta, and South Tangerang. Central Jakarta has the smallest land areas, with most plots clustering between 50 to 150 square meters and a few outliers extending up to 200 square meters, reflecting its dense urban environment and limited space availability. South Tangerang, positioned on the outskirts of Jakarta, has the largest land areas among the regions, with a median around 200 square meters and several plots extending well beyond 400 square meters, including significant outliers reaching up to 600 square meters. This distribution reflects the suburban nature of South Tangerang, where larger plots are more common due to less dense development. Overall, the boxplot reveals that land availability and plot sizes vary significantly, with Central Jakarta constrained by limited space, while South Tangerang offers the most expansive plots, aligning with its suburban character.

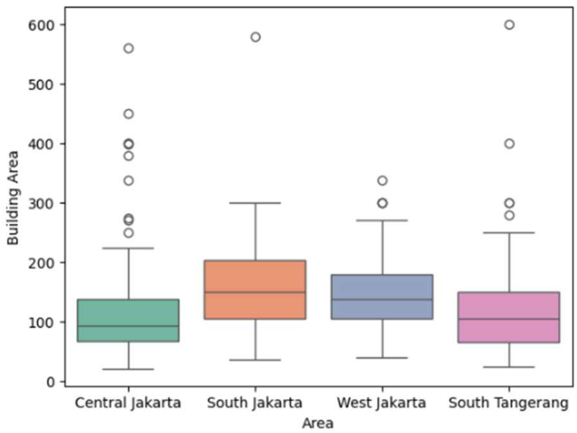

Figure 7. Building Area Boxplot

Figure 7 shows the distribution of building areas across Central Jakarta, South Jakarta (South J), West Jakarta, and South Tangerang. South Jakarta stands out with the largest building areas, having a median size around 200 square meters, with most building sizes ranging between 150 and 300 square meters, and several outliers reaching beyond 400 square meters. This reflects South Jakarta's status as a high-demand area with more spacious residential buildings, aligning with its larger land areas and higher property prices. Central Jakarta has the smallest building areas, with a median size of about 100 square meters and a range that generally stays below 150 square meters, indicative of the compact and dense urban environment where smaller buildings are the norm due to space constraints. The building area distribution is closely linked to land availability, economic factors, and regional characteristics, with South Jakarta and South Tangerang providing contrasting examples of high-density versus spacious suburban development.

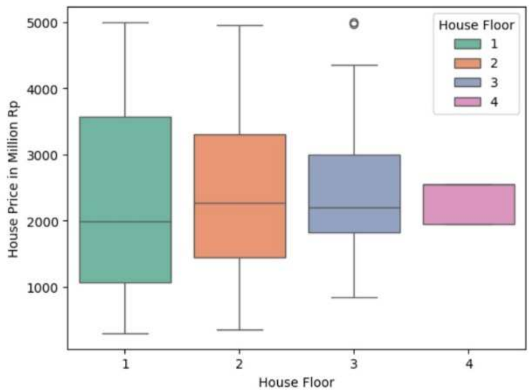

Figure 8. House Floor Boxplot

The boxplot in **Figure 8** displays the distribution of house prices based on the number of floors. Houses with a single floor have the highest median price, along with a wider range of prices, which suggests that these houses often cover larger areas and, as a result, are priced higher. In contrast, houses with two, three, and four floors tend to have lower median prices with narrower interquartile ranges, indicating more consistent pricing. The data implies that the number of floors might be inversely related to the price due to the smaller area of multi-floor houses compared to large, single-floor homes.

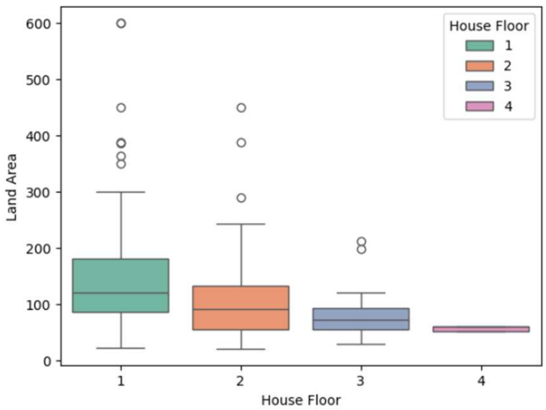

Figure 9. House Floor and Land Area Boxplot

The boxplot in **Figure 9** shows the relationship between house floors and land area. Houses with a single floor tend to occupy larger land areas, as indicated by the higher median and the wide range of values, including many outliers. This supports the idea that single-floor houses are often built on larger plots, which may explain why they don't require additional floors. In contrast, houses with more floors tend to occupy smaller land areas, with less variation in their size. This suggests that multi-floor houses may be built on smaller plots, leading to a more compact structure.

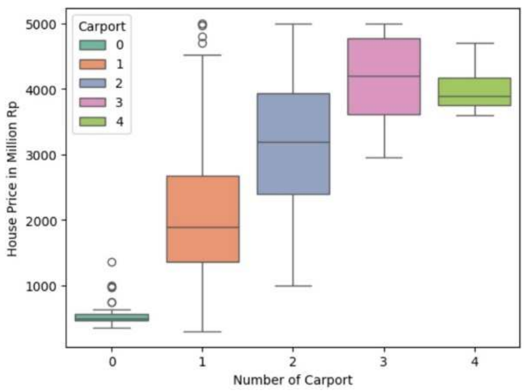

Figure 10. Carport Boxplot

The box plot in **Figure 10** shows a clear trend where the average house price increases as the number of carports increases. Houses without a carport typically have prices below 1 billion rupiah. When a house has one carport, the average price rises to around 2 billion rupiah. As the number of carports increases further, the house prices continue to rise, with houses that have four carports reaching prices close to or exceeding 4 billion rupiah. This data suggests a significant positive correlation between the number of carports and the overall price of a house, indicating that more carports are associated with higher house prices.

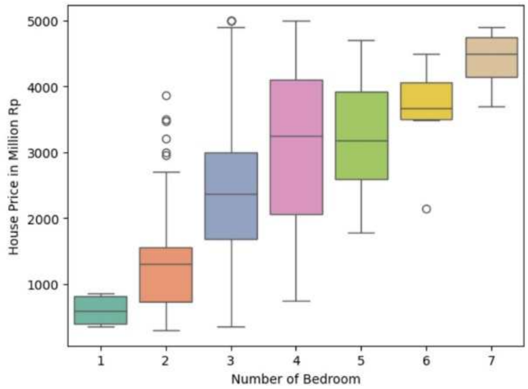

Figure 11. Bedroom Boxplot

The box plot in Figure 11 illustrates how house prices vary based on the number of bedrooms. The data reveals that houses with more bedrooms tend to have higher prices. Specifically, houses with 2 to 3 bedrooms generally fall within the 1.5 to 2.4 billion rupiah range. As the number of bedrooms increases, the price distribution broadens, with houses having 4 to 7 bedrooms showing a wider range of prices and higher median values. This indicates a positive correlation between the number of bedrooms and house prices, with larger homes commanding higher prices.

3.2 Regression Modeling

Obtained using the grid search k-folds cross validation method with k=5 in the three models with the following indicators

Table 2. Tuning Parameter

Method	Tuning Parameter
Multiple Linear Regression	No Tuning
SVR	Kernel, C, Epsilon
Random Forest	n Estimator, Max Depth, Min Samples Split, Min Samples Leaf, Max Features
LGBM	Num Leaves, Learning Rate, n Estimators

The grid search with k-folds cross-validation (k=5) was used to fine-tune the parameters for three models, as outlined in Table 2, with parameters specific to each method, such as kernel and regularization for SVR, and tree-based hyperparameters for Random Forest and LGBM. Once the optimal parameters were identified, the models were tested on a separate dataset that had not been used during training or validation to assess their predictive performance on unseen data. Before interpreting the modelling, we will be interpreting first about descriptive analysis of the housing data which have several key trends related to carports, bedrooms, regional differences, and house levels that influence house prices. Furthermore, descriptive analysis showed that:

1. The more carports there are, the more expensive the house price, with the average house having one carport and the average price being 1.9 billion.
2. The bedrooms were also the same as the average house having two to three bedrooms, with an average of 1.4 billion and 2.5 billion.
3. Based on the region, South Jakarta is the area with the most expensive average of around 2.6 billion, followed by a larger average building and land area.
4. The last variable is the house level, which generally has two levels. However, there are some houses that have 1 level floor but have very large land for the size of a house in urban areas, which makes the range of box plots on one level very varied.

After descriptive analysis, we find try to modelling data with k-fold then predicting with others data which not train or testing before. To decide the best modeling results, we used the R^2 and MSE parameter to compare the methods. The results for the R^2 and MSE parameter are as follows:

Table 3. R^2 and MSE Results for Each Algorithm

Method	Testing Result		Prediction Result	
	R^2	MSE	R^2	MSE
MLR	0,58	0,43	0,866	0,13
LGBM	0,8	0,212	0,85	0,15
SVR	0,78	0,23	0,82	0,182
RF	0,8	0,2137	0,864	0,13

It can be seen that the R^2 and also low MSE results tend not to change significantly between the methods with good parameter tuning, namely SVR, random forest, and LGBM, which indicates that the model is quite good at predicting house prices with the available predictor variables. However, the default method, namely Multiple Linear Regression, only produces $R^{2,}$ which is far behind, namely 58%, which means that the parameter tuning carried out in each method works well and can improve the suitability of the model. We also predicting other houses which are not included in the model fold before and get good result for all prediction result. So, the best model from the testing and predicting result are LGBM and random forest. For MLR even though have the better R^2 and MSE in predicting, it just causes by the lack of sample for predicting and not describe the performance model.

Table 4. Variable Importance

Variable	Importance
Bedroom	0.112
Bathroom	0.055
Land Area	0.315
Building Area	0.369
Floor Level	0.027
Carport	0.068
West Jakarta (Encoding)	0.011
Central Jakarta (Encoding)	0.011
South Jakarta (Encoding)	0.001
South Tangerang (Encoding)	0.018

The importance scores suggest that the most influential factors in determining the target variable are Building Area (0.369) and Land Area (0.315), indicating that property size is a key determinant and also Bedroom which have (0.112) while factors like Carport (0.068), Bathroom (0.055), and Floor Level (0.027) play a lesser role. The geographical location, represented by Encoding variables for West Jakarta, Central Jakarta, South Jakarta, and South Tangerang, has minimal impact, with all these locations having very low importance scores, South Jakarta which are almost negligible. It also related to boxplot which the range of price for every region its almost same, the difference is just for the feature from house.

The best parameters after tuning for three machine learning models are as follows: For the Support Vector Machine (SVM) with RBF Kernel, the optimal parameters are C: 1, epsilon: 0.1, and kernel: 'rbf', indicating a balanced trade-off between training and testing errors with a modest tolerance for errors and a kernel suitable for non-linear data. The Gradient Boosting (likely LightGBM) model uses a learning_rate of 0.005, n_estimators: 10,000, and num_leaves: 3, reflecting a very gradual learning process with a high number of iterations and a simple tree structure to avoid overfitting. Lastly, the Random Forest model is tuned with max_depth: 10, max_features: 'sqrt', min_samples_leaf: 1, min_samples_split: 5, and n_estimators: 100, striking a balance between depth, feature selection, and sample requirements to form a robust ensemble with 100 trees

3.3 Discussion

After the modelling and getting the better model with the most importance variable, we can conclude that property size and number of bedrooms which the most considering features before the buyer going to buy house since those features the most determinant price of a house. House prices are increasingly soaring because of speculation about higher resale prices. This is exacerbated by the progressive land property tax that has not yet been implemented in Indonesia. Research by [18] from the Ministry of Finance shows that the implementation of the Rural and Urban Land and Building Tax (PBB P2) can reduce the excessive use of empty land. One of the efforts to reduce the burden of house purchase prices carried out by the market is the existence of a property rental mechanism such as an annual house, boarding house, or apartment rental. Research shows that the ratio of rental prices to house prices in 18 OECD countries is around 4-6% and fluctuates [19]. This means that someone who rents out property will expect to return their capital in the next 16-25 years along with the additional price of the rental property. The ratio of rental prices to house prices at 4-6% also occurs in Indonesia. Apartment occupancy since the pandemic has decreased drastically and has not yet returned to the optimal price range. The main factors affecting apartment rental prices use variable importance calculations. The variable importance was used to determine the importance of the predictor variable in influencing the response. The greater the VIM value, the greater the importance of a variable [20]. The results obtained showed that apartment area was the highest variable with a value of 0.62, followed by area with a score of 0.18, and number of bedrooms with 0.12. This is in line with the areas of Central Jakarta and South Jakarta, which are the destinations for economic activity and have higher prices than West Jakarta and South Tangerang. The average apartment rental price is:

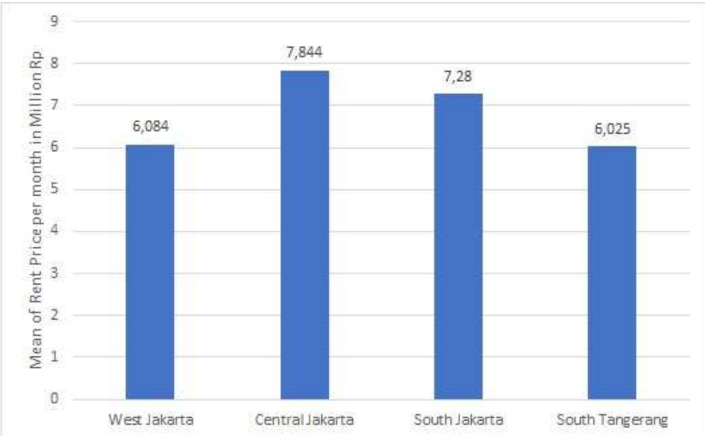

Figure 12. Average Apartment Rental Prices

Figure 12 illustrates the average monthly apartment rental prices across four regions: Central Jakarta, South Jakarta, West Jakarta, and South Tangerang. The data reveals that Central Jakarta has the highest average rental price at 7.844 million, followed closely by South Jakarta at 7.28 million. In contrast, West Jakarta and South Tangerang have lower average rental prices of 6.084 million and 6.025 million, respectively. This trend aligns with the economic activities concentrated in Central and South Jakarta, making these areas more desirable and thus more expensive. The figure supports the broader context of the study, indicating that while apartment rentals in these regions are significantly more affordable than purchasing a house outright, the choice of location remains a critical factor in determining rental prices, with proximity to the city center and economic hubs playing a crucial role.

Apartment rental prices in the 4 areas are still much cheaper than the average price of a house purchased in these areas which is purchased with cash. This conclusion indicates that the advantages of apartments include areas in the middle of the city and cheaper prices because they are vertical residences. On the other hand, in terms of buying and selling small houses and including property in general, many families experience a trade-off between having to choose an expensive property with a strategic location and choosing a cheap one in a suburban area [21] while the mode transport for commuting daily still using private modes due to flexibility and accessibility [22]. The price of apartment rentals includes close proximity, easy access to work, and choice of transportation modes [23]. Other research shows that location factors, number of beds, and proximity to public facilities are also things that renters pay attention to [24].

4. CONCLUSIONS

The study examines the intricate dynamics of housing and apartment markets within Jakarta's metropolitan region, emphasizing the factors influencing property prices and the effectiveness of various predictive modelling approaches. The analysis reveals that key determinants of house prices include the size of the property (land and building area) and the number of bedrooms, with these features significantly impacting price house. While speculative activities and the lack of progressive property taxation exacerbate rising house prices, the research highlights the potential of rental mechanisms as a feasible alternative for alleviating the financial burden on buyers.

The application of machine learning methods, including Multiple Linear Regression (MLR), Support Vector Regression (SVR), Light Gradient Boosting Machine (LGBM), and Random Forest, demonstrates varying degrees of accuracy in predicting house prices. The findings indicate that LGBM and Random Forest models outperform MLR and SVR in terms of R^2 and Mean Squared Error (MSE) which have 0.8 in R^2 and 0.21 for MSE, suggesting their superiority in handling complex datasets and making accurate predictions. The study also underscores the importance of feature selection in model performance, with property size emerging as the most influential variable followed by number of bedrooms.

Additionally, the research exploring into the apartment rental market, where rental prices in central locations like Central and South Jakarta are notably higher due to their proximity to economic area. The analysis shows that while apartment rentals are generally more affordable than purchasing a house, strategic location remains a crucial factor in rental price determination. This finding can be highlights a trade-off faced by families between choosing expensive, strategically located properties and more affordable options in suburban areas.

In conclusion, the study provides valuable insights for prospective homebuyers, investors, and policymakers, emphasizing the importance of property size and location in real estate decisions. It also calls for enhanced property tax policies and a balanced approach to urban development that considers both affordability and accessibility. We also suggest for further analysis to use panel data or considering spatial analysis for comprehensive understanding.

REFERENCES

[1] A. Chmielewska, M. Ciski, and M. Renigier-Biłozor, “Residential real estate investors’ motives under pandemic conditions,” *Cities*, vol. 128, no. May, 2022, doi: 10.1016/j.cities.2022.103801.

[2] M. D. Idawicaksakti, M. D. Itaratnasari, and R. A. Rahadi, “Public Policy and Financial Regulation in Housing Sector (Case Study: One Million Houses and KPR FLPP),” *Int. J. Innov. Enterp. Syst.*, vol. 6, no. 01, pp. 50–60, 2022, doi: 10.25124/ijies.v6i01.127.

[3] D. Hartono, R. A. Budiman, and S. H. Hastuti, “Housing Tenure Choice of Low-Income Household in Jabodetabek,” *Econ. Dev. Anal. J.*, vol. 9, no. 1, pp. 76–86, 2020, doi: 10.15294/edaj.v9i1.37503.

[4] S. Lee, “House prices, homeownership, and household consumption: Evidence from household panel data in Korea,” *Econ. Model.*, vol. 126, no. May, p. 106355, 2023, doi: 10.1016/j.econmod.2023.106355.

[5] D. Sola Gratia and A. Danu Prasetyo, “Determining Factors Influencing the Price of Housing for First Time Home Buyer in Indonesia,” *Int. J. Accounting, Financ. Bus.*, no. 6, pp. 60–73, 2021.

[6] Simon. Zainal Zawir, “APARTEMEN DI JABODETABEK Abstrak,” *J. Orientasi Bisnis dan Interprenersh.*, vol. 1, no. 1, pp. 1–13, 2020.

[7] T. Tjahjono, A. Kusuma, and A. Septiawan, “The Greater Jakarta Area Commuters Travelling Pattern,” *Transp. Res. Procedia*, vol. 47, no. 2019, pp. 585–592, 2020, doi: 10.1016/j.trpro.2020.03.135.

[8] R. A. Rahadi, A. R. Qastharin, R. Bekti, W. Aryakusuma, A. Rahmawaty, and S. P. Groda, “Value Determinant Factors for Apartment Products in Indonesia,” *Rev. Integr. Bus. Econ. Res.*, vol. 9, no. 1, pp. 46–61, 2020.

[9] N. Rizki Novani, N. Rizqihandari2, and M. H. Dewi Susilowati3, “Commuter Housing Arrangement in Beji Sub-District, Depok City, West Java,” *Malaysian J. Sustain. Environ. Spec. Issue*, pp. 107–123, 2022, doi: 10.24191/myse.v9i3.18293.

[10] D. J. Crosss Sihombing, D. C. Othernima, J. Manurung, and J. R. Sagala, “Comparative Models of Price Estimation Using Multiple Linear Regression and Random Forest Methods,” *ICCoSITE 2023 - Int. Conf. Comput. Sci. Inf. Technol. Eng. Digit. Transform. Strateg. Facing VUCA TUNA Era*, pp. 478–483, 2023, doi: 10.1109/ICCoSITE57641.2023.10127705.

[11] D. M. Eler, D. Grosa, I. Pola, R. Garcia, R. Correia, and J. Teixeira, “Analysis of document pre-processing effects in text and opinion mining,” *Inf.*, vol. 9, no. 4, pp. 1–13, 2018, doi: 10.3390/info9040100.

[12] D. Maulud and AM Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," Journal of Applied Science and Technology Trends, vol. 1, no. 4, pp. 140–147, December 2020, doi:https://doi.org/10.38094/jastt1457.

[13] N. K. S, N. V. S, and N. R. R, “A comparative analysis on linear regression and support vector regression,” *2016 Online International Conference on Green Engineering and Technologies (IC-GET)*, doi: 10.1109/get.2016.7916627.

[14] W. Bakasa and S. Viriri, “Light Gradient-Boosting Machine edge detection with cropping layer for semantic segmentation of pancreas,” *Advances in Artificial Intelligence and Machine Learning*, vol. 03, no. 03, pp. 1274–1294, Jan. 2023, doi: 10.54364/aaiml.2023.1175.

[15] Y. Li *et al.*, “Random forest regression for online capacity estimation of lithium-ion batteries,” *Applied Energy*, vol. 232, pp. 197–210, Dec. 2018, doi: 10.1016/j.apenergy.2018.09.182.

[16] E. T. Wahyuningtyas and D. A. Susesti, “Comparison of House Price Analysis in DKI Jakarta and Surabaya with Minitab Software 17.0,” *Procedia Business and Financial Technology*, vol. 1, Aug. 2021, doi: 10.47494/pbft.2021.1.16.

[17] K. E. Mukti, “The Influence of Housing Attributes on Housing Price in East Surabaya,” *2nd International Conference on Indonesian Economy and Development (ICIED 2017)*, Jan. 2018, doi: 10.2991/icied-17.2018.16.

[18] E. S. Ayatuna, “Pegenaan pajak progresif atas persediaan tanah kosong (idle land bank),” *Simposium Nasional Keuangan Negara 2 (1)*. pp. 1014–1115, 2020. [Online]. Available: https://jurnal.bppk.kemenkeu.go.id/snkn/article/view/603

[19] T. Engsted and T. Q. Pedersen, “Predicting returns and rent growth in the housing market using the rent-price ratio: Evidence from the OECD countries,” *J. Int. Money Financ.*, vol. 53, pp. 257–275, 2015, doi: 10.1016/j.jimonfin.2015.02.001.

[20] H. Liang, Z. Guo, J. Wu, and Z. Chen, “GDP spatialization in Ningbo City based on NPP/VIIRS night-time light and auxiliary data using random forest regression,” *Adv. Sp. Res.*, vol. 65, no. 1, pp. 481–493, 2020, doi: 10.1016/j.asr.2019.09.035.

[21] Y. Dewita, M. Burke, and B. T. H. Yen, “The relationship between transport, housing and urban form: Affordability of transport and housing in Indonesia,” *Case Stud. Transp. Policy*, vol. 8, no. 1, pp. 252–262, 2020, doi: 10.1016/j.cstp.2019.01.004.

[22] A. F. Aritenang, “Identifying post-suburbanization: The case of the Jakarta metropolitan area (JMA),” *Habitat Int.*, vol. 138, no. January, p. 102857, 2023, doi: 10.1016/j.habitatint.2023.102857.

[23] J. J. Lin and Y. C. Cheng, “Access to jobs and apartment rents,” *J. Transp. Geogr.*, vol. 55, pp. 121–128, 2016, doi: 10.1016/j.jtrangeo.2016.07.012.

[24] I. D. Amenyah and E. A. Fletcher, “Factors Determining Residential Rental Prices,” *Asian Econ. Financ. Rev.*, vol. 3, no. 1, pp. 39–50, 2013.