Available Online at www.e-iph.co.uk
Indexed in Clarivate Analytics WoS

E-B
Environment - Behaviour
Proceedings Journal

AicE-Bs2025London

https://www.amerabra.org/
13th Asia-Pacific International Conference on Environment-Behaviour Studies
University of Westminster, London, UK, 29-31 Aug 2025

Commercial Shop Price Assessment: Comparing Machine Learning and Ordinary Least Squares for improved accuracy

Junainah Mohamad[1*], Intan Faiqah Hamizah Mohd Firazan[1], Suraya Masrom[2], Abdul Rehman Gilal[3]
**Corresponding Author*

[1] Department of Built Environment Studies and Technology, Faculty of Built Environment, Universiti Teknologi MARA, Perak Branch, Seri Iskandar Campus, Malaysia
[2] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, Malaysia
[3] Knight Foundation School of Computing and Information Sciences, Florida International University, USA

mjunainah@uitm.edu.my, intanfaiqahhamizah@gmail.com, suray078@uitm.edu.my, arehman@fiu.edu
Tel: +60127123044

Abstract
This study compares the predictive performance of OLS and five ML algorithms in valuing commercial shop properties using 2,480 transactions from Kuala Lumpur from 2013 to 2023. While OLS showed limited predictive power, the Random Forest algorithm, applied with log-transformed target variables, achieved superior accuracy (R^2 = 0.9974, RMSE = 0.03, MAPE = 0.02%). These findings support the use of machine learning as a reliable and efficient alternative for property valuation, offering enhanced precision and scalability in commercial real estate assessment.

Keywords: Commercial Valuation; Machine Learning; Random Forest; Ordinary Least Squares

DOI: https://doi.org/10.21834/e-bpj.v10i33.7256

1.0 Introduction

Accurate valuation of commercial properties is fundamental to the functioning of real estate markets, as it shapes investment decisions, guides urban planning, and informs policymaking (Malpezzi, 2003). In rapidly growing cities such as Kuala Lumpur, Malaysia, the expansion of commercial activity and rising demand for retail space intensify the need for robust and reliable valuation models that can reflect market realities (Topraklı, 2025; Khamis et al., 2020). Commercial shops, in particular, represent a vital segment of the urban property market, directly influencing business development, municipal revenue, and the broader urban economy. Thus, precise and data-driven valuation of these assets is critical to sustaining balanced urban growth and investor confidence.

Traditionally, the Ordinary Least Squares (OLS) regression model, also known as the Hedonic Pricing Model (HPM), has been the dominant tool for property valuation, explaining property prices based on structural, locational, and neighbourhood attributes (Abidoye & Chan, 2018). While OLS is valued for its simplicity and interpretability, it is limited by sensitivity to functional form assumptions, multicollinearity, and the inability to capture nonlinear interactions common in complex real estate datasets (Selim, 2009; Bourassa et al., 2025). While hedonic models can be specified in linear, semi-log, or log-log forms, the choice of the most suitable functional form often presents methodological challenges (Owusu-Ansah, 2018), reducing forecasting accuracy and consistency.

In response to these limitations, Machine Learning (ML) methods have emerged as powerful alternatives, representing what Breiman (2001) describes as the *algorithmic modelling culture*. Algorithms such as Decision Trees, Random Forests, Support Vector Regression, and XGBoost can capture nonlinear relationships, high-dimensional interactions, and complex patterns often overlooked by OLS

DOI: https://doi.org/10.21834/e-bpj.v10i33.7256

(Antipov & Pokryshevskaya, 2012). Empirical evidence suggests that ML models frequently achieve higher predictive accuracy, as measured by R^2 and RMSE, though they may sacrifice interpretability compared to traditional regression (Johnson et al., 2023).

This study aims to enhance the accuracy of commercial shop price valuation in Kuala Lumpur by comparing the predictive performance of ML algorithms with OLS regression. Specifically, the study evaluates five machine learning models against OLS using a dataset of 2,480 shop transactions from 2013 to 2023. The objectives of this study are 1)To analyse model performance by comparing the predictive accuracy of OLS and selected ML algorithms using statistical metrics (R^2, adjusted R^2, RMSE, and MAPE) and, 2) To evaluate model outcomes by assessing how well OLS and ML predicted prices align with actual commercial shop transaction prices, thereby identifying their strengths and limitations in practical valuation.

By addressing the lack of direct empirical comparisons between ML and OLS in the valuation of commercial properties, this study contributes to the development of a more robust, scalable, and context-sensitive valuation framework. The findings are expected to support property professionals in enhancing appraisal accuracy and to aid urban policymakers in making evidence-based decision for sustainable city development.

2.0 Literature Review

2.1 Operational Definition of Shop

A shop, also referred to as a shophouse or shop lot, is defined as a building primarily intended for commercial use. The ground floor is typically reserved for retail or business, while the upper floors may be used for offices, residences, or storage. Shops may be standalone units, in shop-office rows, or within mixed-use developments. According to the *Manual Definisi NAPIC* (2021), shops are categorised into five types: pre-war, terrace, semi-detached, detached, and multi-storey shop units or retail lots.

This study covers all sales transactions for these types, excluding strata shop units in shopping complexes. Only shops with express commercial land use conditions are included. These properties may serve multiple functions, including business, residential, storage, institutional, or showroom, either singly or in combination. Physically, Malaysian shops range from single-storey to six-and-a-half storeys (Jamaludin et al., 2021). Sales data reflect whole-unit transactions only.

In practice, commercial shops are commonly appraised using the sales comparison, investment, or cost approach. However, heterogeneity in location, design, and use complicates assessment, highlighting the need for more data-driven approaches.

2.2 Ordinary Least Squares (OLS) Versus Machine Learning (ML)

The HPM, typically estimated with OLS, has long been applied to property valuation (Lucas, 1975; Rosen, 1974). OLS is valued for simplicity but assumes linearity, independence, and normally distributed errors. These assumptions often fail in complex datasets. Multicollinearity among predictors such as land area and location may lead to unstable estimates (Ismail, 2006). Remedies such as variable removal risk losing important information. Furthermore, the choice of linear, semi-log, or log-log forms remains a methodological challenge (Owusu-Ansah, 2018).

ML models such as Decision Trees, Random Forests, Support Vector Regression, and XGBoost overcome these issues by capturing nonlinear relationships and high-dimensional interactions (Yin et al., 2021). While less interpretable, ML offers stronger adaptability and predictive accuracy.

2.3 Gap in Knowledge and Research Contribution

Most valuation studies emphasise housing, with limited focus on commercial shops despite their urban importance. OLS and ML are often applied separately, with few direct comparisons on the same dataset. Appraisal principles are also rarely integrated into modelling frameworks.

This study fills these gaps through a comparative analysis of OLS and ML in predicting Kuala Lumpur shop prices, embedding appraisal principles, and assessing performance using R^2, adjusted R^2, RMSE, and MAPE. Findings aim to improve valuation accuracy and support evidence-based policymaking.

3.0 Research Methodology

This research focuses on commercial shop transactions within the Federal Territory of Kuala Lumpur, which is divided into three main regional clusters under the Malaysian Property Services Index (MSPI): KL Centre, KL North, and KL South. The KL Centre covers Sections 1–100, Mukim Kuala Lumpur, and Mukim Ampang. KL North includes Mukim Batu, Mukim Setapak, and Mukim Hulu Klang. Meanwhile, KL South comprises Mukim Petaling and Mukim Cheras. These regional divisions ensure comprehensive coverage of urban, suburban, and mixed-use areas across Kuala Lumpur, providing a representative dataset for analysis. The section presents the study's process flow, divided into four main phases: data collection and preparation, data analysis, model development and evaluation, implementation, and conclusion. Figure 1 depicts the overall flow.

3.1 Data Collection and Data Preparation

The dataset comprises 2,980 commercial shop property transactions in Kuala Lumpur from 2013 to 2023, sourced from the National Property Information Centre (NAPIC). Only arm-length transactions were retained to ensure accurate price representation. Key variables include transaction details as shown in Table 1.

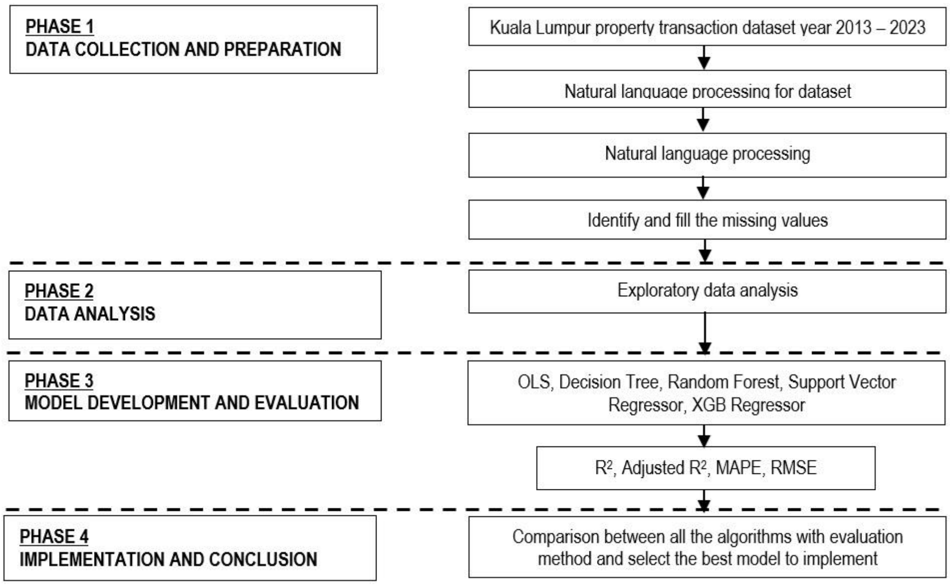

Figure 1. Research Flow Chart
(Source: Researcher)

Table 1. Features described in the dataset

		Variable	Description of variable	Level of measurement
Physical	1	Land Area	The total size of the land on which the shop is located.	Scale – measured in square meters.
	2	Main floor area (MFA)	The total floor area of the main floor of the shop.	Scale – measured in square meters.
	3	Ancillary floor area (AFA)	The floor area of the additional or supporting spaces of the shop, such as walkways.	Scale – measured in square meters.
	4	Main floor Area + ½Ancilary floor area	The total MFA + ½ total; of AFA	Scale – measured in square meters.
	5	Type of Shop	The category or classification of the shop (e.g., terrace shop, detached shop, and semi-detached shop)	Nominal
	6	Position	Position of the shop. (e.g., corner, middle, end and others)	Nominal
	7	Condition of the shop	The overall physical state of the shop. (e.g., new, very good, good, average, poor, bad)	Nominal
	8	Type of Shop Construction	The classification of the shop's construction based on its durability and materials (e.g., permanent, semi-permanent, and temporary)	Nominal
	9	Number of Floors	The total number of floors in the shop or building. (e.g., 1,2,3,4)	Ratio – measured in number according to the number of floors
	10	Terrain	The characteristics or classification of the land surface on which a property is built or located.	Nominal
Location	11	Region	The region of Kuala Lumpur is divided by 3 areas Central, North, and South.	Nominal
	12	Mukim	The subdistrict or division within a district where the property is located.	Nominal
	13	Scheme	The specific development scheme under which the shop is located	Nominal
	14	Street layer	The type or classification of the road layer the shop is situated in, such as first, second, third, and inner layer.	Nominal
	15	Classification of location	The broader classification of the location includes primary city centre, secondary city centre, primary rural area, and secondary rural area.	Nominal
	16	Condition of the location	The overall condition or state of the location, such as good, average, or poor.	Nominal
	17	Land use	The designated use of the land includes building, industrial, agriculture, and no categories.	Nominal
	18	Distance to nearest city	The distance of the shop to the nearest city benchmark is usually the main post office.	Scale – measure by kilometre
Legal	19	Type of Lot	The legal classification or designation of the land based on ownership rights or registration details.	Nominal
	20	Tenure	The type of property tenure, indicating whether the shop is freehold or leasehold property,	Nominal

21	Purchaser Status	The citizenship or type of entity purchasing the shop.	Nominal
22	First Transfer	Indicates whether the shop is being transferred for the first time or sub-sale.	Nominal

(Source: Researcher)

Data cleaning involved removing records with missing or inconsistent values, correcting errors, handling outliers identified through Z-score analysis (Chikodili *et al.,2020)*, and encoding categorical variables using one-hot encoding. The filtering process reduced the dataset to 2,480 transactions (Table 2). Continuous variables were normalised to a standard scale to facilitate model convergence, and categorical variables were encoded appropriately.

Table 2. Record of data cleaning process

No.	Notes	Total deleted records	Number of records left
1.	Original data from 2013-2023 for commercial shop property (1-4 story shop)	-	2980
2.	Missing land area	2	2978
3.	Missing MFA	2	2976
4.	Missing shop type of construction	64	2912
5.	Missing terrain	80	2832
6.	Missing street layer	180	2652
7.	Missing condition of location	121	2531
8.	Outliers	51	**2480**

(Source: Researcher)

3.2 Data Analysis

Exploratory data analysis provided statistical summaries and insights into the distributions, relationships, and potential multicollinearity of variables. Table 3 details the descriptive statistics of the data.

Table 3. Descriptive statistics for the final dataset

Variables	Mean	Median	Mode	Std. Dev
Region	1.92	2.00	1	.813
Mukim	7.40	2.00	2	16.220
Scheme	503.46	480.00	360	309.834
Land Area	262.7094	153.300	164.00	1993.52290
MFA	428.3817	434.1500	515.20	179.72882
AFA	29.9484	22.2900	.00	23.82530
MFAAFA	443.3575	446.4000	528.00	183.06718
Type of Shop	1.00	1.00	1	.020
Position	3.05	4.00	4	1.049
Condition	3.11	3.00	3	.876
Type Construction	1.01	1.00	1	.087
Number of Floor	2.88	3.00	3	.798
Terrain	8.55	9.00	9	1.989
Street Layer	1.70	2.00	1	.879
Classification of Location	2.12	2.00	2	.861
Condition of Location	1.64	2.00	1	.685
Land Use	1.06	1.00	1	.412
Distance to Nearest City (km)	11.0195	12.0000	15.00	4.69337
Tenure	.45	.00	0	.498
Purchaser Status	5.34	8.00	8	2.943
First Transfer	.05	.00	0	.223
Type of Lot	1.27	1.00	1	.687
Price	2475617.79	2159000.00	2000000	1305439.084
PriceLog	14.588262	14.585000	14.5100	.5335532
PricePSM	5851.9842	5316.0950	4339.35	2900.84912
PricePSMLog	8.582153	8.580000	8.6800	.4187696

(Source: Researcher)

The correlation matrix (Figure 2) revealed a strong relationship between MFA and the number of floors. The data was removed from the database for model development. Multicollinearity diagnostics using the Variance Inflation Factor (VIF) confirmed acceptable levels (VIF < 5).

Figure 3 is typical of real estate or sales price distributions, where many items are priced relatively low, but a small number of high-priced items skew the average upward. The mean price is 2,475,617.79 (RM). This is the average price across the dataset. The standard deviation is 1,305,439.084 (RM), indicating a significant spread in prices, as some prices deviate substantially from the mean. The histogram is right-skewed, with most of the data clustered at the lower price ranges. The highest frequency is for prices below 2,000,000 (RM), and the frequency decreases sharply as prices increase. The long tail to the right suggests the distribution is positively skewed, meaning that most properties or items are priced below the mean, but some very high prices pull the mean higher than the median.

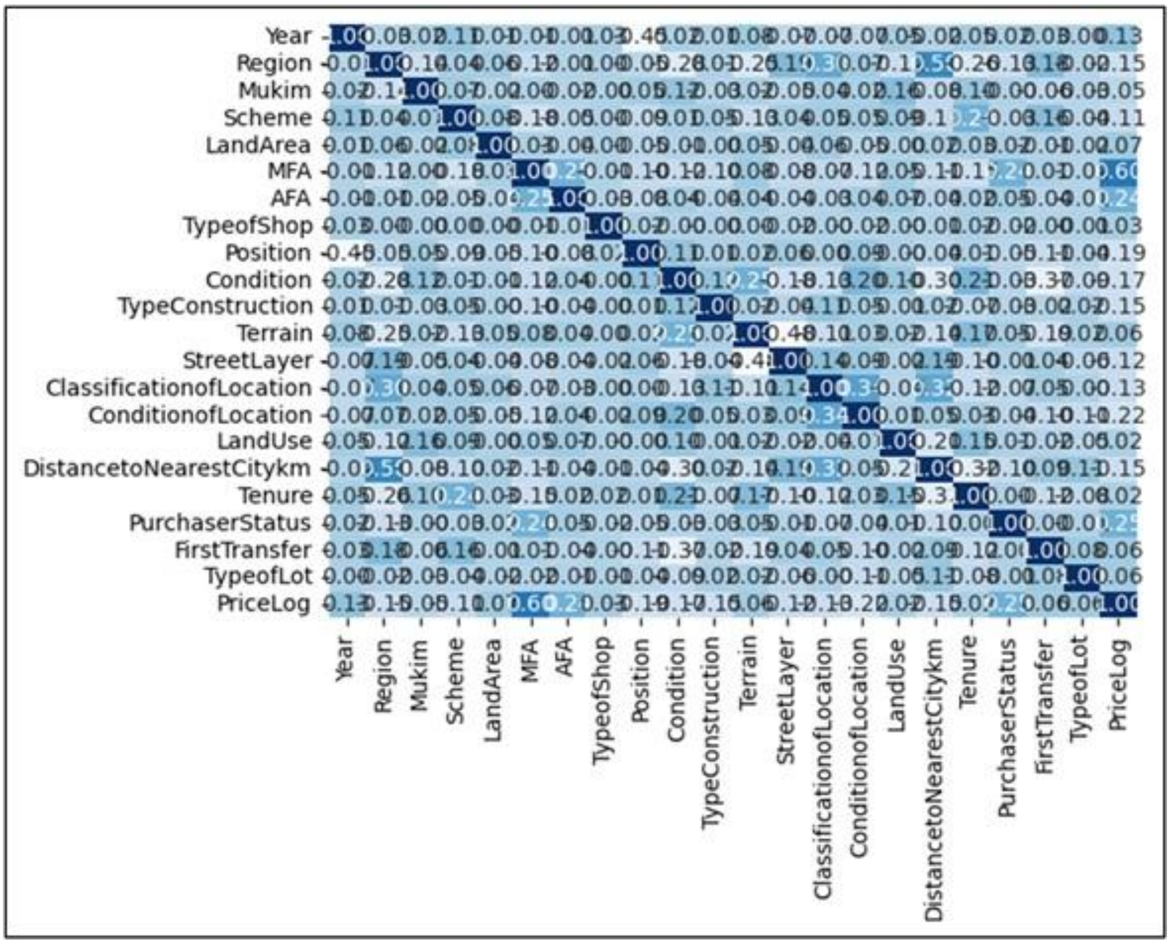

Figure 2. Correlation matrix of the dataset
(Source: Researcher)

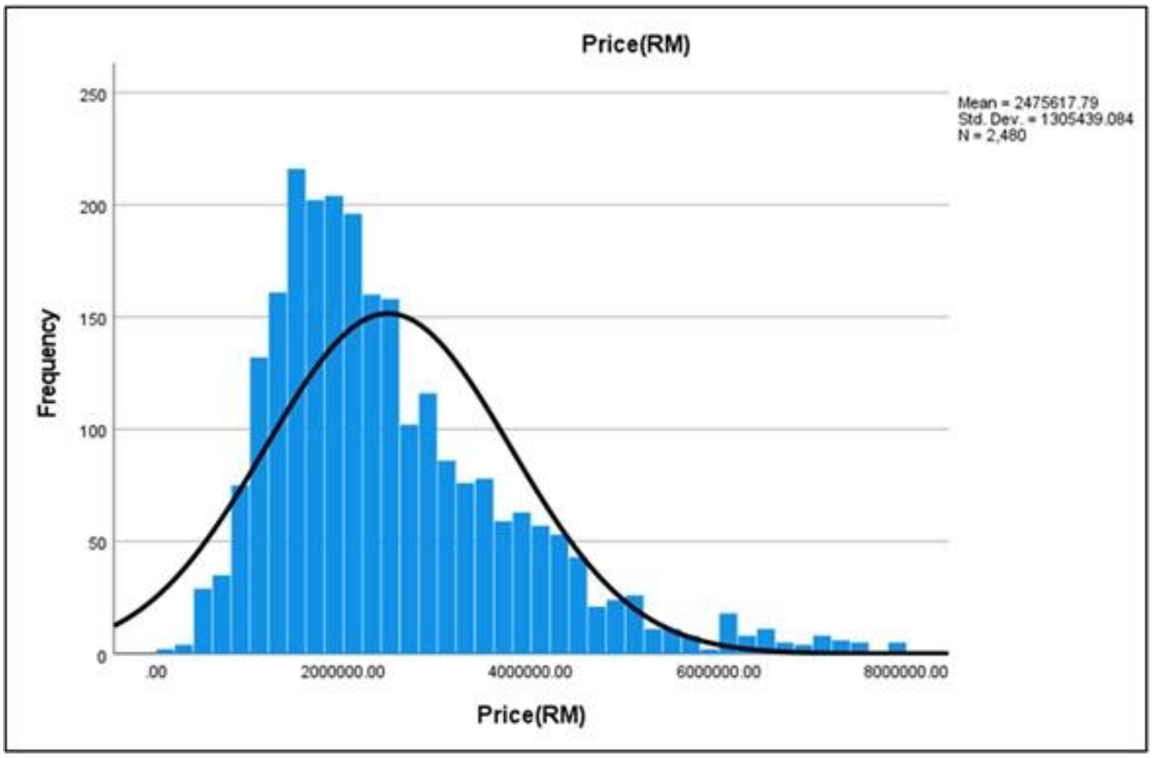

Figure 3. Histogram
(Source: Researcher)

3.3 Model Development and Evaluation

OLS regression was conducted using the *Enter* method in Statistical Package for the Social Sciences (SPSS). The statistical significance of the predictors was assessed based on *t*-values, with a threshold of $|t| > 2$ indicating significance.

Five ML models were developed and tested: Decision Tree Regressor, Random Forest Regressor, Support Vector Regressor, XGBoost Regressor, and MLP Regressor. These models were implemented in Python using Google Colab. A stratified 70:30 train-test split was employed to ensure representative and robust model validation. Hyperparameter tuning was carried out using grid search in combination with cross-validation to optimise each model's performance.

The performance of all models was assessed using four key metrics: Coefficient of Determination (R^2), Adjusted R^2, Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). Additionally, logarithmic transformations of the target variables (i.e., *PriceLog*) were applied to stabilise variance and potentially enhance predictive accuracy.

4.0 Results and Discussion

In this section, the results of OLS and ML algorithms are presented and discussed, with a focus on selecting the best model for commercial shop properties.

4.1 Findings of Objective 1: Model Performance

Table 4 presents a comparative analysis of OLS and five ML algorithms for predicting commercial shop prices. Results show that Decision Tree and Random Forest outperformed all other models, delivering significantly higher predictive accuracy and lower error rates.

The Decision Tree achieved the best overall performance, with an R^2 of 0.9984, RMSE of 0.02, and MAPE of 0.01%. Random Forest followed closely, with $R^2 = 0.9974$, RMSE = 0.03, and MAPE = 0.02%. These models effectively captured nonlinear relationships and demonstrated strong model fit. In contrast, OLS regression performed poorly ($R^2 = 0.452$), confirming its limitations in modelling complex real estate data.

Other models, such as XGBoost, showed moderate accuracy, while SVR and MLP regressors performed poorly, with negative R^2 values and MAPE above 5%, indicating high prediction errors and poor generalisation.

Overall, tree-based models, particularly Decision Tree and Random Forest, emerged as the most effective for shop price prediction. The use of log-transformed target variables further enhanced prediction stability by reducing variance.

In the next phase, a comparative analysis will be conducted between Random Forest and Decision Tree using actual versus predicted price plots. This will help determine which model produces predictions that more closely align with actual market values, offering further insight into their reliability for commercial valuation tasks.

Table 4. Descriptive statistics for the final dataset

No.	Model/Algorithms	R^2	Adjusted R^2	RMSE	MAPE
1.	OLS	.452	.449	-	-
2.	**Decision Tree**	**0.9984**	**0.9984**	**0.02**	**0.01%**
3.	**Random Forest**	**0.9974**	**0.9974**	**0.03**	**0.02%**
4.	Support Vector Regressor	-1.4311	-1.4344	0.83	5.45%
5.	XGBoost	0.9465	0.9465	0.12	0.06%
6.	MLP Regressor	-3.4329	-3.4389	1.12	6.07%

(Source: Researcher)

4.2 Findings of Objective 2: Model Evaluation: Actual Price Vs Predicted Price

This section assesses six regression models: OLS, Decision Tree, Random Forest, Support Vector Regressor (SVR), XGBoost, and MLP Regressor in predicting log-transformed commercial shop prices in Kuala Lumpur, based on 2,480 transactions from 2013 to 2023. Performance was evaluated using histogram overlays (Figure 4) and scatter plots of actual vs predicted prices (Figure 5).

Among all models, the Random Forest Regressor delivered the most consistent and accurate results. Its prediction distribution closely mirrors actual values (Figure 4), and its scatter plot (Figure 5) demonstrates strong alignment along the trend line, with a low RMSE of 62.80. The model's ensemble structure enabled it to capture non-linear patterns while avoiding overfitting.

In contrast, the Decision Tree showed broader dispersion in both figures, with less precise alignment and a higher RMSE of 105.92. Although more straightforward and more interpretable, Decision Trees are more prone to overfitting, particularly with complex, high-dimensional data.

While the MLP Regressor showed tightly clustered predictions, its exceptionally low RMSE (1.316) may indicate overfitting. OLS and SVR performed the weakest, with OLS failing to capture complex variable interactions.

Overall, Random Forest proved to be the most balanced in terms of accuracy, generalisability, and interpretability, making it the most suitable model for commercial shop price estimation in dynamic urban markets.

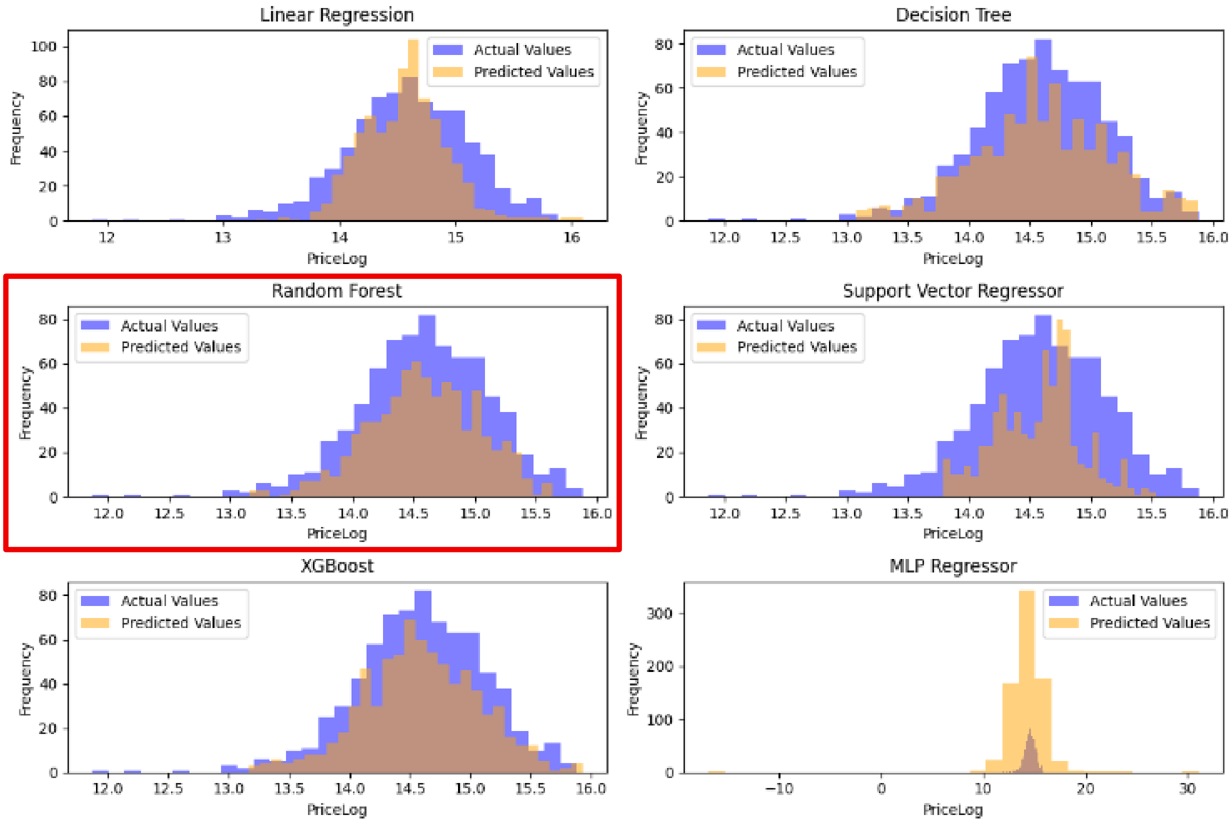

Figure 4. Distribution of Actual and Predicted Price (Kuala Lumpur)
(Source: Researcher)

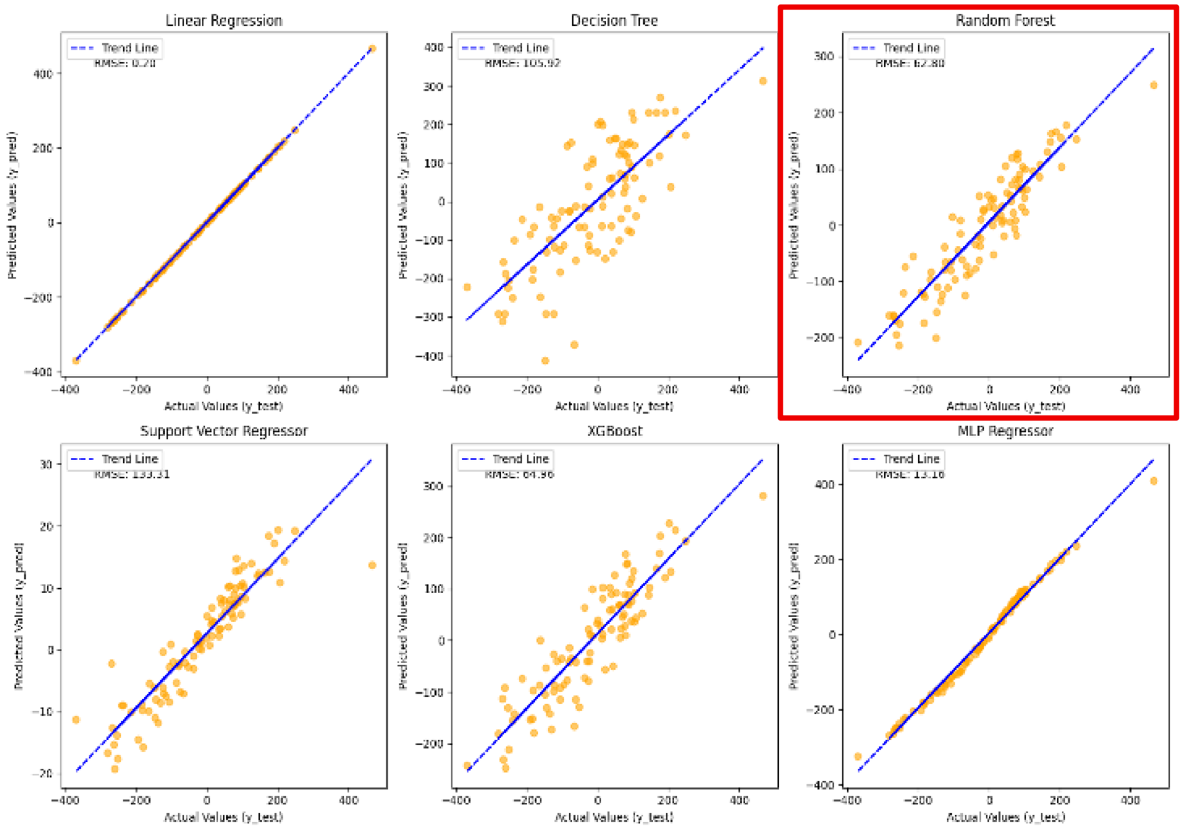

Figure 5: Scatter plot of predicted prices trained with all variables
(Source: Researcher)

5.0 Results and Discussion

This study developed a machine learning-based valuation framework for commercial shop properties in Kuala Lumpur, using 2,480 transactions (2013–2023) and comparing several algorithms with OLS. Models were evaluated based on speed, predictive accuracy, predictor importance, and their ability to capture nonlinear relationships.

Random Forest consistently outperformed other models, achieving an R^2 of 0.9999, MAPE of 0.19%, and RMSE of 0.03. Its ensemble structure captured complex interactions and reduced overfitting, aligning with previous studies (Antipov & Pokryshevskaya, 2012; Yin et al., 2021). While slower than OLS, it provided insights into predictor importance, enhancing transparency. OLS remained beneficial as a benchmark but showed reduced accuracy in high-dimensional data. Decision Tree and XGBoost performed moderately well, while SVR and MLP were less effective. Log-transformation improved prediction stability across all models.

6.0 Conclusion and Recommendations

Random Forest emerged as the most reliable and robust model for predicting shop prices in Kuala Lumpur, outperforming OLS and other algorithms. Its scalability and ability to model non-linearity make it a practical tool for valuation in heterogeneous markets.

It is recommended that valuers and policymakers incorporate ML-based frameworks, particularly Random Forest, into practice to improve appraisal accuracy, support transparent taxation, and strengthen evidence-based urban planning. Future research could extend this framework to other property types, regional comparisons, or integration with geospatial and macroeconomic data.

Acknowledgements

The authors would like to express their heartfelt appreciation to the Ministry of Finance (MOF) Malaysia for supporting this research through the National Real Estate Research Coordinator (NAPREC) Grant Scheme under the 12th Malaysia Plan (RMK-12). Special thanks are also extended to the Valuation and Property Services Department (JPPH) for the administrative and technical support associated with the grant [Grant No.: 100-TNCPI/GOV 16/6/2 (050/2024)]. The authors also gratefully acknowledge the insightful comments and constructive feedback from the anonymous reviewers, which greatly improved the quality of this paper.

Paper Contribution to Related Field of Study

This paper advances property valuation by applying machine learning for accurate, scalable commercial price prediction.

References

Abidoye, R. B., & Chan, A. P. (2018). Achieving property valuation accuracy in developing countries: the implication of data source. *International Journal of Housing Markets and Analysis*, *11*(3), 573-585.

Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: An application of Random Forest for valuation. *Expert Systems with Applications*, 39(2), 1772–1778. https://doi.org/10.1016/j.eswa.2011.08.077

Bourassa, S. C., Hoesli, M., Mayer, M., & Stalder, N. (2025). Reflections on hedonic price modeling. *Journal of European Real Estate Research*.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Chikodili, N. B., Abdulmalik, M. D., Abisoye, O. A., & Bashir, S. A. (2020, November). Outlier detection in multivariate time series data using a fusion of K-medoid, standardized euclidean distance and Z-score. In *International Conference on Information and Communication Technology and Applications* (pp. 259-271). Cham: Springer International Publishing.

Chong, W. T., Lim, L. H., & Masron, T. (2020). Hedonic House Price Modelling: Comparing OLS and Machine Learning Methods. *International Journal of Housing Markets and Analysis*, 13(1), 105–123. https://doi.org/10.1108/IJHMA-06-2019-0060

Ismail, S. (2006). Spatial autocorrelation and real estate studies: A literature review. *Malaysian Journal of Real Estate*, *1*(1), 1-13.

Jamaludin, M. H., Ismail, S., & Ismail, N. (2021). The Development of Penang Shop Price Index (PSPI) Using Laspeyres Hedonic Price Model. In *Journal of the Malaysian Institute of Planners VOLUME* (Vol. 19).

Jang, D., & Lee, B. (2022). When machine learning meets social science: A comparative study of ordinary least square, stochastic gradient descent, and support vector regression for exploring the determinants of behavioral intentions to tuberculosis screening. *Asian Communication Research*, *19*(3), 101-118.

Johnson, S., Elms, J., Madhavan, K., Sugasi, K., Sharma, P., Kurban, H., & Dalkilic, M. M. (2023, August). Are They What They Claim: A Comprehensive Study of Ordinary Linear Regression Among the Top Machine Learning Libraries in Python. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Long Beach, CA, USA* (pp. 6-10).

Khamis, R., Mohamed, A. A., & Mohd, R. (2020). Real estate valuation using machine learning: A review. *International Journal of Advanced Science and Technology*, 29(6), 1905–1917.

Lucas, R. E. (1975). Hedonic price functions. *Economic Inquiry*, *13*(2), 157-178.

Malpezzi, S. (2003). Hedonic pricing models: A selective and applied review. In T. O'Sullivan & K. Gibb (Eds.), *Housing Economics and Public Policy* (pp. 67–89). Blackwell Science.

Mayer, M., Bourassa, S. C., Hoesli, M., & Scognamiglio, D. (2019). Estimation and updating methods for hedonic valuation. *Journal of European real estate research*, *12*(1), 134-150.

Owusu-Ansah, A. (2018). *Construction and application of property price indices*. Routledge.

Pusat Maklumat Hartanah Negara (NAPIC). (2021). *Manual Definisi NAPIC* (3.0).

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, *82*(1), 34-55.

Sauerbrei, W., Perperoglou, A., Schmid, M., Abrahamowicz, M., Becher, H., Binder, H., ... & TG2 of the STRATOS initiative Abrahamowicz Michal Becher Heiko Binder Harald Dunkler Daniela Harrell Frank Heinze Georg Perperoglou Aris Rauch Geraldine Royston Patrick Sauerbrei Willi. (2020). State of the art in selection of variables and functional forms in multivariable analysis—outstanding issues. *Diagnostic and prognostic research*, *4*(1), 3.

Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications*, 36(2), 2843–2852. https://doi.org/10.1016/j.eswa.2008.01.044

Topraklı, A. Y. (2025). AI-driven valuation: a new era for real estate appraisal. *Journal of European Real Estate Research*, *18*(1), 105-120.

Yin, L., Recker, W., & Chen, A. (2021). Improving property valuation with machine learning: Evidence from the real estate market. *Computers, Environment and Urban Systems*, 85, 101549. https://doi.org/10.1016/j.compenvurbsys.2020.101549