Journal of Informatics and Web Engineering

Vol. 5 No. 2 (June 2026) eISSN: 2821-370X

Machine Learning-based Prediction of House Sale Prices in Hulu Langat

Kath Moon Yap Choo[1], Sew Lai Ng[2*], An Da[3]**

[1,2]Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Malaysia.

[3]Faculty of Computer and Big Data, Jining Normal University, 59 Gongnong Road, 012000 Ulanqab, China.

**corresponding author: (slng@mmu.edu.my; ORCiD: 0000-0002-8610-581X)*

***corresponding author: (15122453209@163.com; ORCiD: 0009-0006-8733-408X)*

Abstract - House price prediction remains a complex task due to the interplay of structural, locational, and economic factors. This study proposes a machine learning-based predictive framework tailored to the residential real estate market in Hulu Langat, Malaysia, a district undergoing rapid urbanization. Using open-source datasets, three structured variations are developed: one containing only housing and locational attributes, another incorporating regional-level macroeconomic data, and a third combining both regional and national macroeconomic indicators. The proposed model (Ensemble Weighted Average) is trained and evaluated using evaluation metrics, then compared with models such as Random Forest, XGBoost, LightGBM, and Ensemble Stacking. The proposed model trained solely on housing and locational data achieved the highest accuracy, outperforming all other models across most evaluation metrics, while XGBoost achieves the fastest computation time. Models trained with the inclusion of macroeconomic indicators consistently underperforms, suggesting that macroeconomic indicators added noise to model prediction, potentially due to spatial resolution mismatches or multicollinearity. The interpretability of the best-performing model was further enhanced with SHapley Additive exPlanations (SHAP), the resulting SHAP analysis reveals that land parcel area, property type, and local housing supply are the top contributing features to model's performance. These findings validate the effectiveness of ensemble models for localized price prediction and highlight the importance of house attributes over broader economic trends. The proposed framework yields a practical and interpretable approach to house price prediction and may assist policymakers, developers, and planners in making informed decisions.

Keywords—House Price Prediction, Machine Learning, Ensemble Learning, SHAP, Read Estate Analytics, Housing Market.

Received: 19 June 2025; Accepted: 18 August 2025; Published: 16 June 2026

1. INTRODUCTION

The residential real estate market represents the cornerstone for developed and developing economies, as it is one of the most fundamental basic human needs. Malaysia, as an emerging economy, is currently experiencing rapid urbanization, admission districts such as Hulu Langat have seen a significant population growth and increased attention from homebuyers due to its proximity to Kuala Lumpur. However, the determinants of house prices are often complex and influenced by a multidimensional interplay of factors. But recent breakthrough in Machine Learning (ML) technologies and big data analysis enables a new opportunity for deploying an accurate predictive framework

https://doi.org/10.33093/jiwe.2026.5.2.7

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

for localized housing markets. The research aims to utilize advanced ML algorithms to address the unique challenges and opportunities of house price prediction in Hulu Langat, providing a detailed and localized perspective. The findings contribute to the research on ML applications in house price prediction, demonstrate the effectiveness of advanced models, offer a foundational ML framework for future studies in Malaysia, and support policymakers in urban planning strategies.

The complex nature of house price prediction can particularly be explained by theories such as Hedonic Pricing Theory (HPT) [1], Spatial Economic Theory [2], and the law of supply and demand [3]. HPT posits that in a free marketplace, houses are treated as a collection of unique, inseparable property characteristics such as house size [4], [5], [6], basement size [7], and external structure [4]. Spatial Economic Theory further extends by emphasizing that spatial distribution of economic activity near Central Business District (CBT) [7], schools, public transport, and shopping malls [4], and public facilities [8], plays a significantly role in influences house prices, making location a pivotal factor. Moreover, the law of supply and demand states that rising demand can lead to price surges when supply is constrained. Due to these factors, effective integration of these complex, multidimensional, and often non-linear factors into accurate and interpretable predictive models remains a significant challenge for traditional statistical techniques.

The complex and multidimensional factor often hampers traditional models, such as Hedonic model and Multiple Regression Analysis (MRA) [10], [11]. Although MRA can outperform more advanced ML models with smaller datasets [11], advanced ML models generally provide superior performance for complex datasets [10], [11], [12]. However, these advanced models are often affected by issues such as interpretability [12] and overfitting [11], [13]. Notably, hybrid models have demonstrated the capacity to mitigate overfitting while also achieving higher performance than standalone advanced ML models [13].

The research holds significant implications by contributing to the existing body of literature on ML applications in house price prediction, particularly within the underexplored Hulu Langat district. It demonstrates the effectiveness of advanced ML models in house price prediction, while provide a basis ML framework for future studies relating to house price prediction in Hulu Langat context or Malaysian context. Furthermore, the findings support policymakers in decision-making on housing policies and urban planning strategies.

2. LITERATURE REVIEW

2.1 Determinants of the Housing Market

The housing market and house prices are influenced by a myriad of factors, which can be broadly categorized into 2 different overarching subsections, such as internal factors and external factors. This section aims to explore the many factors that have been identified from various existing studies and research that have contributed to the field of house price analysis and prediction.

2.1.1 Internal Factors

Internal factors refer to house attributes such as size, number of rooms, bathrooms, and structural design. These features directly influence buyer preferences and market demand, with their impact varying across local market conditions [14]. In Malaysia, housing satisfaction is closely linked to built-up size, number of bathrooms and rooms, as well as internal and external structural features [4]. House size and number of bathrooms exhibit a strong positive correlation with price, while car parks and rooms have moderate effects. In contrast, greater distances to amenities such as schools and public transport are negatively associated with house prices [6]. In the Ames dataset, general living area and basement size significantly influence prices, while in the Melbourne dataset, number of rooms, bathrooms, land area, and proximity to CBD are the most impactful attributes [7]. In Selangor, attributes such as lot size, tenure, number of bedrooms, and building size are significant determinants, with district-specific variations observed in areas like Petaling Jaya and Shah Alam [5].

2.1.2 External Factors

External factors encompass macroeconomic indicators and geographic proximity. Real GDP accounts for nearly 60% of house price changes in Malaysia and is similarly influential in the USA and OECD countries [15], [16], [17], [18]. Inflation, CPI, and median household income also show strong positive correlations with Malaysian house prices, reflecting economic stability and affordability trends [4], [18], [19], [20]. Although population growth is statistically insignificant in the U.S. context [16], it correlates positively with housing demand in Malaysia [21].

Geographical factors further shape price dynamics. In Kuala Lumpur, shorter distances to schools, MRT/LRT stations, and shopping malls are moderately associated with higher prices, while proximity to hospitals has a weaker effect [6]. Comparable results in Beijing confirm the importance of access to public facilities in house price modelling [8].

2.2. Comparison of ML Models with Traditional Models

Traditional models such as the Hedonic Pricing Model (HPM) are widely used in house price prediction due to their simplicity and interpretability [22]. However, they are limited in handling nonlinear relationships and often underperform in large or complex datasets [10]. With the increasing availability of housing data, ML models have become more prominent for their superior predictive accuracy.

Several studies demonstrate that models like RF, K-Nearest Neighbours (KNN), and Artificial Neural Networks (ANN) consistently outperform traditional models such as HPM [12], [23]. For instance, in Kuala Lumpur, Malaysia, XGBoost and LightGBM achieved higher accuracy than MRA and Ridge Regression [6]. In Taiwan, both regression models and Long Short-Term Memory (LSTM) networks demonstrated strong performance in forecasting the Taiwan House Price Index (THPI) [19], while Back Propagation Neural Networks (BPNN) and Convolutional Neural Networks (CNN) also showed high accuracy [24].

Ensemble methods such as Hybrid Regression Models (e.g., 65% Lasso, 35% XGBoost) and Stacked Generalization have been shown to outperform standalone ML models like RF, XGBoost, and LightGBM [6]. Additionally, the lack of interpretability in complex ML models has been addressed through explainable AI methods such as SHAP, applied to both linear and tree-based models for clearer insight into feature importance [7].

These findings collectively suggest a shift from traditional statistical models to more powerful and flexible ML and deep learning models, particularly when enhanced through hybridization and explainable methods.

3. RESEARCH METHODOLOGY

This section details the research methodology employed for developing the proposed house price prediction framework in Hulu Langat. It details a structured and systematic workflow to design extensive experimentation, model development, model testing, and insightful evaluation of the framework.

3.1 Proposed Hybrid ML Model

The proposed Hybrid ML model utilizes the weighted averaging hybrid approach in combining three distinct models, RF [25], XGBoost [26], and LightGBM [27] to enhance predictive accuracy and robustness by exploiting their complementary strengths. The final prediction of the hybrid model, P_{Hybrid}, is calculated as:

$$P_{Hybrid} = w_1 \cdot P_{RF} + w_2 \cdot P_{XGB} + w_3 \cdot P_{LightGBM}$$

Where:

- P_{RF}: Final prediction of RF.
- P_{XGB}: Final prediction of XGBoost.
- $P_{LightGBM}$: Final prediction of LightGBM

Where w_1, w_2, and w_3 are the respective weights assigned to the RF, XGBoost, and LightGBM predictions, chosen such that $w_1 + w_2 + w_3 = 1$ (Figure 1).

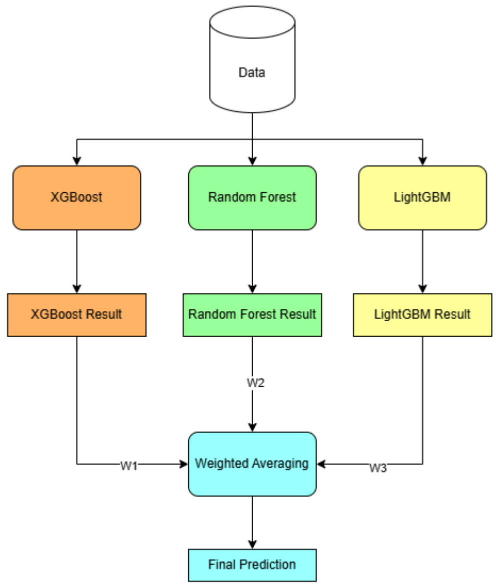

Figure 1. Ensemble Weighted Average Model Diagram

3.2 Research Procedure and Experimental Design

The research methodology employs a structured, five-phase approach, as visualized in Figure 2, to develop and evaluate the house price prediction framework and identify the key determinants of house prices.

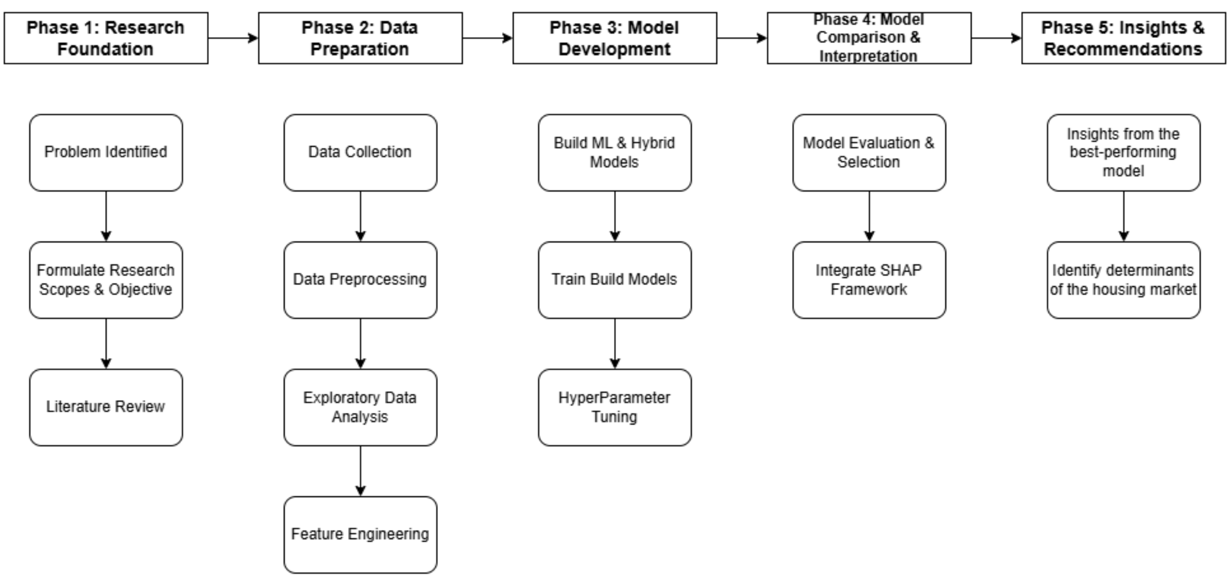

Figure 2. Research Procedure

3.2.1 Phase 1: Research Foundation

The initial phase focused on identifying the core challenges for house price prediction, then formulating the research scopes and research objectives to give the research a clear goal to achieve. Finally, past works and studies done on the field of house price prediction were researched and reviewed to gain better understanding and insight into the field.

3.2.2 Phase 2: Data Preparation

Data collection focused on Malaysian national and regional contexts, specifically for economics and housing. Mainly sources from government platforms such as National Property Information Centre (NAPIC), OpenDOSM, and Department of Statistics Malaysia (DOSM), supplemented third-party platforms such as OSM and Google Maps. The raw data were pre-processed to ensure quality and consistency, including data cleaning, transformation, and integration into a unified dataset.

Exploratory Data Analysis (EDA) was performed to understand data distributions, identify outliers, and examine inter-variable relationships and multicollinearity. Feature engineering included encoding categorical data, normalizing house price using $log1p$, and geocoding is conducted on location data to extract latitude/longitude coordinates and calculate distance measurements to key public facilities. For experimental purposes, three distinct datasets were prepared:

Dataset 1: House attributes, coordinates, and proximity features

Dataset 2: Dataset 1 with regional-level macroeconomic indicators

Dataset3: The full dataset with national-level macroeconomic data.

3.2.3 Phase 3: Model Development

ML models such as RF, XGBoost, and LightGBM were developed alongside hybrid approaches, including a proposed hybrid model and an ensemble stacking model (uses RF and XGBoost as the base learner, and Ridge Regression as the meta learner). These models were selected based on their demonstrated effectiveness in previous literature and their suitability for structured tabular data.

Each model was separately on the training sets of the three distinct datasets to evaluate performance variations across different data compositions. Hyperparameter tuning was performed using both grid search and random search strategies on models that achieved the best cross-validation scores. This process aims to improve prediction accuracy and mitigate the risk of overfitting.

3.2.4 Phase 4: Model Comparison & Interpretation

All models were evaluated on their corresponding test sets with evaluation metrics, including Root Mean Squared Logarithmic Error (RMSLE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), coefficient of determination (R^2), adjusted R^2, and computation time, which is the time taken during model evaluation.

The interpretability best-performing models were enhanced by using the SHAP algorithm. The algorithm produces a Shapley value for each feature input into the best-performing model, which enables further analysis of each feature's contribution to the model's predictions. This allows a clear understanding into the key features influencing the model house price predictions.

3.2.5 Phase 5: Insights & Recommendations

The final phase focused on analysing the resultant Shapley values derived from the best-performing model, to identify the key determinants of house prices in Hulu Langat district. These findings enable stakeholders such as policymakers, developers, and homebuyers to make an informed decision regarding Hulu Langat's housing market, by providing them with valuable information.

4. RESULTS AND DISCUSSIONS

The experimental results are shown in Table 1.

Table 1. Performance Evaluation Scores on All Models Trained on 3 Distinct Datasets

Models	Dataset	RMSLE	RMSE	MAE	MSE	R^2	Adj R^2	Computation Time (s)
Random Forest	1	0.0123	0.1687	0.1156	0.0285	0.9443	0.9441	0.1535
	2	0.0123	0.1690	0.1164	0.0286	0.9442	0.9437	0.1064
	3	0.0126	0.1730	0.1202	0.0300	0.9414	0.9402	0.1247
XGBoost	1	0.0121	0.1655	0.1171	0.0274	0.9464	0.9462	**0.0096**
	2	0.0122	0.1675	0.1187	0.0281	0.9451	0.9447	0.0302
	3	0.0125	0.1718	0.1214	0.0295	0.9423	0.9411	0.0190
LightGBM	1	0.0121	0.1666	0.1180	0.0277	0.9458	0.9455	0.0206
	2	0.0123	0.1688	0.1196	0.0285	0.9443	0.9438	0.0164
	3	0.0124	0.1703	0.1213	0.0290	0.9433	0.9421	0.0170
Ensemble Weighted Average Model	1	**0.0117**	**0.1608**	**0.1119**	**0.0258**	**0.9495**	**0.9493**	0.3293
	2	0.0119	0.1629	0.1140	0.0265	0.9481	0.9477	0.1500
	3	0.0121	0.1667	0.1172	0.0278	0.9457	0.9445	0.1547
Ensemble Stacking Model	1	0.0118	0.1624	0.1138	0.0264	0.9484	0.9482	0.1244
	2	0.0119	0.1635	0.1149	0.0268	0.9477	0.9473	0.1312
	3	0.0122	0.1676	0.1182	0.0281	0.9450	0.9439	0.1474

Experimental results demonstrate that the Ensemble Weighted Average Model trained on Dataset 1 achieved the best overall performance across all evaluation metrics, except for computational time, where XGBoost model trained on Dataset 1 recorded the fastest execution.

The better performance of the ensemble method can be attributed to the ensemble techniques that combines the strengths of 3 base learner models: RF offering robustness to overfitting [25], while XGBoost and LightGBM provide high accuracy through additive optimization [26], [27], compel with the weighted averaging ensemble techniques have shown their ability to improve generalization and reduce model variance in structured data applications [28], [29].

The stark performance gap between models trained on dataset without macroeconomic data and those trained incorporating them. Furthermore, models with macroeconomic input have shown to consistently underperform, suggesting these features added noise rather than predictive value. This may stem from mismatches in spatial granularity (national/state-level data applied to district-level predictions) and multicollinearity among macro indicators [30], [31]. Figure 3 visualizes the generated Shapley values from the best-performing model.

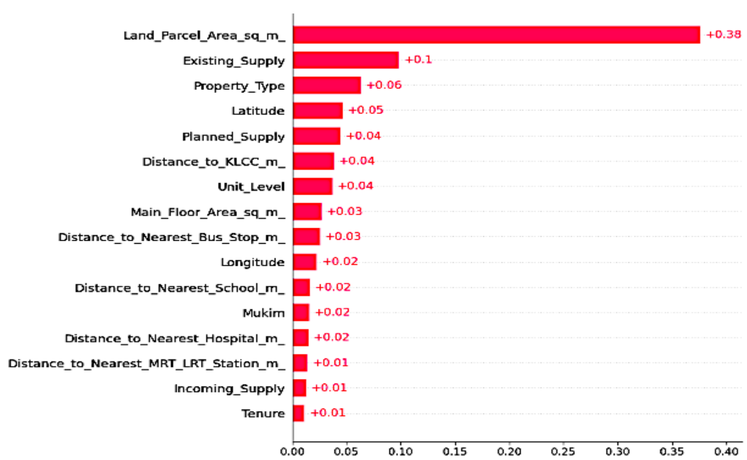

Figure 3. SHAP Summary Plot

Figure 3 reveals that land parcel size, number of existing supplies, and property types are among the top contributors. These findings align with HPT [32] , which posits that property value arises from its attributes. Larger land sizes generally contribute positively to price, reflecting scarcity and development potential. Property type differentiates housing utility and ownership desirability, landed homes command higher premiums than flats. The influence of existing supply suggests localized market dynamics also shape house prices.

5. CONCLUSION

The study presents a ML framework for house price prediction in Hulu Langat, Malaysia, using an ensemble of Random Forest, XGBoost, and LightGBM. Among all tested models, the proposed Ensemble Weighted Average Model trained on dataset of house attributes, coordinates, and proximity features, achieved the highest performance, demonstrating the model's effectiveness in capturing nonlinear relationships in structured tabular data.

Comparative experiments reveal that incorporating macroeconomic indicators do not enhance prediction accuracy; in fact, performance declined slightly across all models. This suggests that localized housing features are more predictive than broader economic variables at the district level.

The SHAP analysis offers valuable interpretability by empirically identifying the most influential features driving house price predictions in Hulu Langat. Key predictors included land parcel area, existing supplies, and property type. These findings support established real estate valuation principles, particularly HPT, which emphasizes the role of intrinsic property attributes and location.

Future works should further enhance the capabilities of the proposed system and methodology, by gathering a larger and more diverse localized dataset across time, improving geographical data quality, and experimenting with different ML or deep learning models such as ANN Model, Joint Self Attention Model, CatBoost Model, and Long Short-Term Memory Model that uses the same or different ensemble techniques, and combinations, to further enhance the predictive capabilities.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their suggestions to improve the paper.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Kath Moon Yap Choo: Conceptualization, Data Curation, Methodology, Validation, Visualization, Writing – Original Draft Preparation;
Sew Lai Ng: Project Administration, Review & Editing;
Da An: Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests was disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] J. Montero, and G. Fernandez-Aviles, "Hedonic price model", in Encyclopedia of Quality of Life and Well-Being Research, Springer, Dordrecht, pp. 2834-2837, 2014, doi: 10.1007/978-94-007-0753-5_1279.

[2] K. Satoru, "Overview of major models of spatial economics," IDE Discussion Paper, 2024. [Online]. Available: https://www.ide.go.jp/English/Research/Topics/Eco/Spatial/overview.html

[3] D. Gale, "The law of supply and demand," *Mathematica Scandinavica*, vol. 3, no. 1, pp. 155–169, 1955, doi: 10.7146/math.scand.a-10436.

[4] N. Amit, H. Sapiri, and Z. Md Yusof, "Factors affecting housing price in Malaysia using structural equation modeling approach," *Sains Malaysiana*, vol. 51, no. 12, pp. 4161-4173, 2022, doi: 10.17576/jsm-2022-5112-23.

[5] Y. F. Chang, W. C. Choong, S. Y. Looi, W. Y. Pan, and H. L. Goh, "Analysis of housing prices in Petaling district, Malaysia using functional relationship model," *International Journal of Housing Markets and Analysis*, vol. 12, no. 5, pp. 884–905, 2019, doi: 10.1108/ijhma-12-2018-0099.

[6] S. Abdul-Rahman, N. H. Zulkifley, I. Ismail, and S. Mutalib, "Advanced machine learning algorithms for house price prediction: Case study in Kuala Lumpur," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 12, 2021, doi: 10.14569/ijacsa.2021.0121291.

[7] E. Z. Teoh, W.-C. Yau, T. S. Ong, and T. Connie, "Explainable housing price prediction with determinant analysis," *International Journal of Housing Markets and Analysis*, vol. 16, no. 5, pp. 1021–1045, 2022, doi: 10.1108/ijhma-02-2022-0025.

[8] P.-Y. Wang, C.-T. Chen, J.-W. Su, T.-Y. Wang, and S.-H. Huang, "Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism," *IEEE Access*, vol. 9, pp. 55244–55259, 2021, doi: 10.1109/ACCESS.2021.3071306.

[9] G. C. Chow, and L. Niu, "Housing prices in urban china as determined by demand and supply," *Pacific Economic Review*, vol. 20, no. 1, pp. 1–16, 2015, doi: 10.1111/1468-0106.12080.

[10] S. N. Abd. Rahman, N. H. Adi Maimun, M. N. Mohamed Razali, and S. Ismail, "The Artificial Neural Network model (ANN) for Malaysian housing market analysis," *Planning Malaysia Journal*, vol. 17, no. 9, May 2019, doi: 10.21837/pm.v17i9.581.

[11] N. Nguyen and A. Cripps, "Predicting housing value: A Comparison of multiple regression analysis and artificial neural networks," *Journal of Real Estate Research*, vol. 22, no. 3, pp. 313–336, 2001, doi: 10.1080/10835547.2001.12091068.

[12] M. Yazdani, "Machine learning, deep learning, and hedonic methods for real estate price prediction," *ArXiv Preprint ArXiv211007151*, 2021.

[13] Q. Truong, M. Nguyen, H. Dang, and B. Mei, "Housing price prediction via improved machine learning techniques", *Procedia Computer Science*, vol. 174, pp. 433-442, 2020, doi: 10.1016/j.procs.2020.06.111.

[14] S. C. Bourassa, D. R. Haurin, J. L. Haurin, M. Hoesli, and J. Sun, "House price changes and idiosyncratic risk: The impact of property characteristics," *Real Estate Economics*, vol. 37, no. 2, pp. 259–278, 2009, doi: 10.1111/j.1540-6229.2009.00242.x

[15] S. H. Kok, N. W. Ismail, and C. Lee, "The sources of house price changes in Malaysia," *International Journal of Housing Markets and Analysis*, vol. 11, no. 2, pp. 335–355, 2018, doi: 10.1108/IJHMA-04-2017-0039.

[16] X. Ding, "Macroeconomic factors affecting housing prices: Take the United States as an example," in *Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)*, Atlantis Press, pp. 2335–2339, 2022, doi: 10.2991/aebmr.k.220307.380.

[17] U. A. Hassan Fereidouni Gholipour, and A. H. Mohammed, "Foreign investments in real estate, economic growth and property prices: evidence from OECD countries," *Journal of Economic Policy Reform*, vol. 17, no. 1, pp. 33–45, 2014, doi: 10.1080/17487870.2013.828613.

[18] S. H. Zulkarnain, A. S. Nawi, M. A. Esquivias, and A. Husin, "Determinants of housing prices: evidence from East Coast Malaysia," *International Journal of Housing Markets and Analysis*, Emerald Group Publishing Limited, vol. 18, No. 3, pp. 573-597, January. 2024, doi: 10.1108/IJHMA-10-2023-0139.

[19] K.-C. Chiu, "A long short-term memory model for forecasting housing prices in Taiwan in the post-epidemic era through big data analytics," *Asia Pacific Management Review*, vol. 29, no. 3, pp. 273–283, 2024, doi: 10.1016/j.apmrv.2023.08.002.

[20] S. Pillaiyan, "Macroeconomic drivers of house prices in Malaysia," *Canadian Social Science*, vol. 11, no. 9, pp. 119–130, 2015, doi: 10.3968/7482.

[21] T. San Ong, "Factors affecting the price of housing in Malaysia," *Journal of Emerging Issues in Economics, Finance and Banking*, vol. 1, pp. 414–429, 2013.

[22] A. Owusu-Ansah, "A review of hedonic pricing models in housing research," Journal of International Real Estate and Construction Studies, vol. 1, no. 1, pp. 19-38, 2011.

[23] V. Limsombunchao, "House price prediction: hedonic price model vs. artificial neural network," 2004.

[24] C. Zhan, Z. Wu, Y. Liu, Z. Xie, and W. Chen, "Housing prices prediction with deep learning: An application for the real estate market in Taiwan," in *2020 IEEE 18th International Conference on Industrial Informatics (INDIN)*, pp. 719–724, 2020, doi: 10.1109/INDIN45582.2020.9442244.

[25] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/a:1010933404324.

[26] T. Chen, and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[27] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, vol. 30, 2017.

[28] Z.-H. Zhou, "Ensemble methods: foundations and algorithms," *CRC press,* 2012, doi: 10.1201/b12207.

[29] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, pp. 1–39, 2009, doi: 10.1007/s10462-009-9124-7.

[30] D. E. Farrar and R. R. Glauber, "Multicollinearity in regression analysis: The problem revisited," *The Review of Economics and Statistics*, vol. 49, pp. 92–107, 1967, doi: 10.2307/1937887.

[31] E. L. Glaeser and J. Gyourko, "The impact of building restrictions on housing affordability," *Economic Policy Review, Federal Reserve Bank of New York*, issue Jun, pp. 21–39, 2003.

[32] S. Rosen, "Hedonic prices and implicit markets: Product differentiation in pure competition," Journal of Political Economy, vol. 82, no. 1, pp. 34–55, 1974, doi: 10.1086/260169.

BIOGRAPHIES OF AUTHORS

	**Kath Moon Yap Choo** is a Bachelor candidate in Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia. His research focuses on finance informatics, machine learning and the application of predictive modeling in finance. His specialization includes feature engineering, geocoding, advanced modeling, explainable AI and business intelligence. He can be contacted at email: yapkathmoon@gmail.com.

	**Sew Lai Ng** is an Assistant Professor in the Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia. Her research interests lie in the development and application of predictive modeling, statistical and econometric techniques, time series analysis, and risk management. She focuses on leveraging advanced modeling approaches to understand complex systems and inform decision-making in various domains. She can be contacted at email: slng@mmu.edu.my.
	**Da An** is a lecturer at the School of Computer and Big Data, China Jining Normal University. He obtained his Bachelor of Engineering degree from China Tianjin Polytechnic University in 2016 and his Master of Engineering degree from Tianjin Polytechnic University in 2019. His research interests include Image Processing, Deep Learning, and Robotics Manufacturing. He can be contacted at email: 15122453209@163.com.