International Journal for Computational Civil and Structural Engineering, 20(3) 69–83 (2024)

DOI:10.22337/2587-9618-2024-20-3-69-83

AN INNOVATIVE HYBRID MACHINE LEARNING TECHNIQUES FOR PREDICTING CONSTRUCTION COST ESTIMATES

Kittisak Lathong, Kittipol Wisaeng

Mahasarakham Business School, Mahasarakham University, Mahasarakham, THAILAND

Abstract: The importance of precise cost estimations in construction investment sector cannot be overstated, serving as the cornerstone for crucial investment decisions. Unfortunately, inaccurate estimations often lead to significant losses due to flawed investment objectives. Consequently, the pursuit of an effective investment decision support system has become a paramount research focus, aiming to aid construction investors in making informed and timely choices. This study focuses on a novel hybrid machine learning (HML) approach, amalgamating various base models such as artificial neural networks (ANNs), support vector machines (SVMs), multiple linear regression (MLR), decision trees (DTs) and random forest (RF) to construct a sophisticated construction cost forecasting model. Remarkably, empirical findings demonstrate the exceptional accuracy of the hybrid ANN-DT model, reaching an impressive 92.1% and surpassing individual models. This breakthrough holds the promise of substantial advantages and increased profitability within the construction industry, particularly benefitting professionals in civil engineering, architecture, and construction investing. By combining the predictive strength of ANNs with the transparent decision rules of DTs, this hybrid model effectively addresses the industry's need for precise predictions and understandable forecasting methodologies, representing a significant advancement in enhancing informed decision-making processes within the construction domain.

Keywords: construction cost estimation, machine learning, hybrid learning model, residential buildings

ИННОВАЦИОННЫЕ ГИБРИДНЫЕ МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ СТОИМОСТИ СТРОИТЕЛЬСТВА

К. Латонг, К. Висенг

Бизнес-школа Махасаракхам, Университет Махасаракхам, Махасаракхам, ТАИЛАНД

Аннотация: Сложно переоценить важность точной оценки затрат в сфере инвестиций в строительстве, поскольку это служит краеугольным камнем для принятия важнейших инвестиционных решений. К сожалению, неточные оценки часто приводят к значительным убыткам из-за неверных инвестиционных целей. Поэтому создание эффективной системы поддержки принятия инвестиционных решений стало первостепенным направлением исследований, призванных помочь инвесторам в строительство сделать обоснованный и своевременный выбор. Данное исследование посвящено новому гибридному подходу машинного обучения (HML), объединяющему различные базовые модели, такие как искусственные нейронные сети (ANN), машины векторов поддержки (SVM), множественная линейная регрессия (MLR), деревья решений (DT) и случайный лес (RF), для построения сложной модели прогнозирования стоимости строительства. Примечательно, что эмпирические результаты демонстрируют исключительную точность гибридной модели ANN-DT, достигающую впечатляющих 92,1% и превосходящую отдельные модели. Это обещает существенные преимущества и повышение рентабельности в строительной отрасли, что будет полезным для специалистов в области гражданского строительства, архитектуры и инвестирования в строительство. Сочетая предсказательную силу ANN с прозрачными правилами принятия решений DT, эта гибридная модель эффективно удовлетворяет потребности отрасли в точных прогнозах и понятных методологиях прогнозирования, представляя собой значительное достижение в улучшении процессов принятия обоснованных решений в строительной сфере.

Ключевые слова: оценка стоимости строительства, машинное обучение, гибридная модель обучения, жилые здания

1. INTRODUCTION

Precisely estimating the construction cost of residential buildings through machine learning harnesses the real estate landscape of Thailand. The importance of residential living or housing structures within the flat-building paradigm is crucial to human life, representing a fundamental aspect of human shelter. The burgeoning real estate sector in Thailand has seen a significant surge in construction, notably following the COVID-19 situation that impacted the country and the globe, resulting in economic slowdowns. However, the current market dynamics present challenges in accurately appraising property values. In this context, the assessment process of property valuation is transitioning from traditional methodologies to predictive machine learning techniques.

The cost estimation method [1] serves as a fundamental cornerstone in engineering endeavors associated with entire projects. It exerts significant influence across various aspects, encompassing planning, project feasibility studies, bidding, design, construction management, and cost control. The outcomes of such estimations play a crucial role in assisting project planners and stakeholders in evaluating project feasibility and efficiently managing costs within the intricate scope of detailed project design. Given the inherent scarcity of data available for comprehensive project coverage in the initial phases, construction managers rely extensively on their experience to conduct these estimations. This practice greatly facilitates accurate project cost estimation, offering enhanced convenience in project cost evaluations [2]. Comprehensive cost estimation entails complex assessments involving materials, overhead costs, labor quantities, space, public utilities, sales considerations, temporary parameters, and arrays of associated costs all encapsulated within the project's structural framework and timeline.

Preliminary or rough construction cost estimation entails an initial approximation of project expenses, even as the finer details of the components remain incompletely delineated. The accuracy of this preliminary estimate is reliant upon the estimator's expertise and may incorporate pertinent data extrapolated from analogous completed undertakings. This estimation process commonly manifests a range of variability, typically spanning between 10% and 30% [3], [4], as depicted in Fig. 1. However, Kawee Wangnivejankul [4] posits that this variability may potentially extend up to 50%, thereby introducing an elevated level of risk into the construction undertaking. Consequently, prudent consideration should be accorded to potentially circumventing the preliminary estimation process when feasible, given its propensity to exert a significant influence on the inherent risk profile associated with the construction endeavor. The process of detailed construction cost estimation becomes tenable when equipped with a comprehensive and exhaustive design schema. This entails an intricate dissection of construction materials into quantifiable units, subsequently affording the quantification of costs spanning materials, labor, machinery, operations, profit margins, tax implications, and financial interests, among other factors.

Empirical cost models rely on statistical analysis and historical data curve-fitting, predominantly using capacity as the primary variable. Concurrently, cost estimation by area or volume entails computing project expenses based on the dimensions of a structure or material, encompassing surface area (in square footage) or volume (in cubic meters) considerations. Area-based estimation involves multiplying the cost per unit area by the total area, while volume-based estimation multiplies the cost per unit volume by the total required volume. These widely employed methods in construction facilitate precise budgeting and planning for various development projects, particularly for materials like concrete, paint, and roofing. Parametric models leverage multiple variables, often employing empirical or hybrid empirical/factored approaches, providing tailored specificity for different applications compared to typical parametric or factored cost

models. Factored cost models, commonly utilized in water treatment and oil and gas sectors, rely on capital cost estimates for major equipment and incorporate factors for the remaining capital costs, usually requiring design development and vendor quotations. Material take-off, employed post-significant design activity, involves tallying components and furnishing material schedules, along with cost quotations for each line item. While highly accurate, this method necessitates a well-developed design [5]. Empirical and parametric models are predominantly favored in the early stages of development, with the Association for the Advancement of Cost Engineering (AACE) [6] offering guidelines on estimate accuracy at various project stages. Figure 1 outlines estimate uncertainty and the typical level of design detail provided.

The utilization of advanced machine learning and regression techniques enhances the capability to predict property values of flat-building structures, depending on influential factors within diverse market domains. This study explores innovative decision tree algorithm models, which are unsupervised learning paradigms that function without stringent parameter dependence. These algorithms serve for both classification and regression analysis, synthesizing the core principles of decision-making algorithms [7] derived from the inherent nature of data. The overarching objective is to construct an efficient model capable of expertly predicting target variable values. Machine learning (ML) has found widespread applications across various industrial and business sectors. ML technology has undergone extensive development, expanding predictive capabilities in diverse applications. The intriguing study illustrates the efficacy of ML methods in forecasting or classifying various factors, such as interest rates and real estate prices within specified contexts. Construction cost estimation stands as a pivotal issue under scrutiny, featuring a comprehensive and intricately detailed examination of ML capabilities. Furthermore, construction cost projections pose nonlinear challenges influenced by various construction years and features, both direct and indirect.

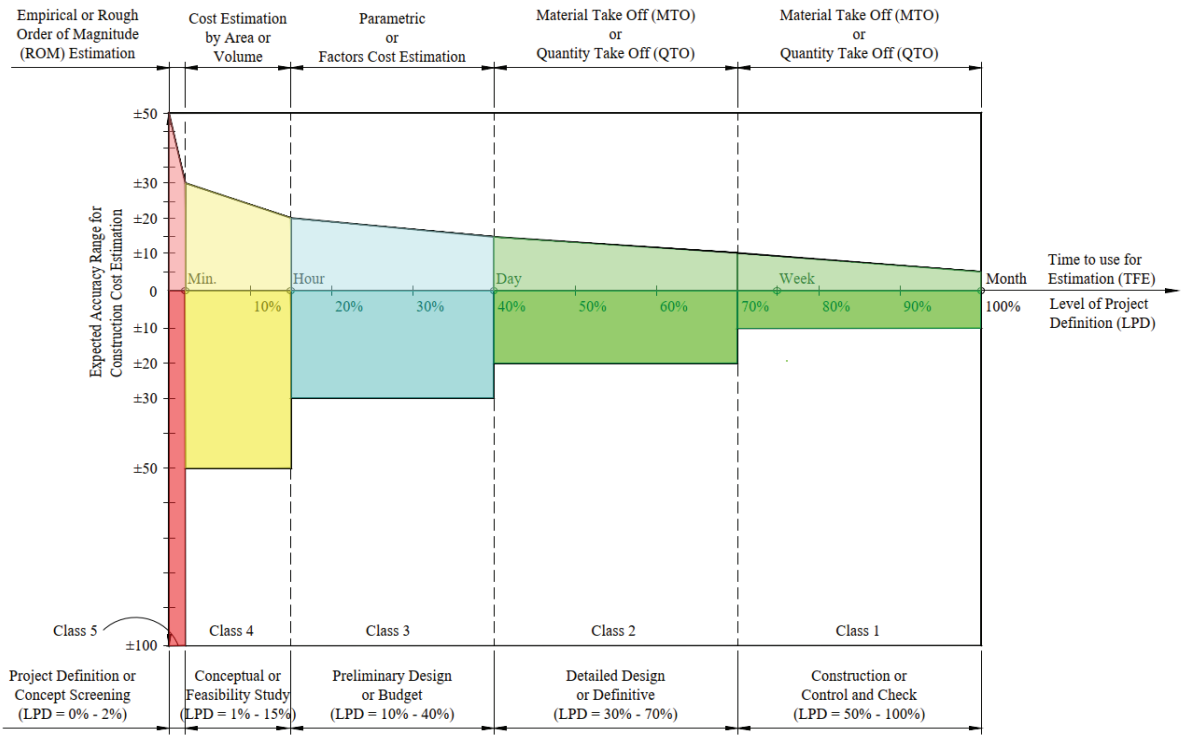

Figure 1. The relationship between construction cost estimate accuracy, the time allocated for estimation, and the project's developmental stage

Machine learning (ML) is a subset of artificial intelligence (AI) that collaborates with algorithms and technologies to extract valuable insights from data. ML is suitable for calculations associated with data because it cannot process data efficiently and independently without ML. Hence, the reliance on algorithms enabling ML to be a predictive algorithm capable of quantitatively processing data for predictions, as stated by Xu et al. [8]. ML has significantly transformed various industries and emerged as a powerful tool in the construction sector, automating processes. ML technology is crucial for processing vast amounts of data, saving time, and enhancing processing resource efficiency. Moreover, it may be particularly suitable for the construction industry in predicting financial costs and time, achieving optimal efficiency [9].

Within the realm of construction cost estimation, Tayefeh Hashemi et al. [10] conducted an in-depth analysis of research papers over a span of 30 years, from 1985 to 2020. These papers aimed to apply ML techniques for cost estimation in construction projects. The overarching goal of these studies was to develop predictive models capable of accurately estimating costs, especially before bidding, facilitating informed decision-making by project managers based on project data. Notably, widely used ML techniques have been employed in reviewed literature, including Artificial Neural Networks (ANN), Regression Analysis (RA), Case-Based Reasoning (CBR), and Support Vector Machines (SVM). This analysis aligns with the findings of Elfaki et al. [11] in their survey of construction cost estimation over the past decade, emphasizing the enduring prominence of classic ML techniques, particularly ANN and SVM, within the field.

The several research studies have illuminated the complexity and significance inherent in predicting real estate prices, acknowledging the diverse factors that wield influence within the housing market. Sifei Lu et al. [12] underscored the necessity for refined methodologies extending beyond the conventional House Price Index. Their emphasis lay in accurately predicting individual house prices, considering a spectrum of factors like location, house type, size, construction year, and local amenities. They proposed a hybrid Lasso and Gradient Boosting regression model, showcased by its exceptional performance in the Kaggle Challenge "House Prices: Advanced Regression Techniques," ranking among the top 1% among all participants, signifying promising predictive capabilities.

Sruthi Chiramel et al. [13] recognized the challenges faced by homebuyers due to inflation and escalating housing prices. Their study aimed to predict house prices in Iowa, USA, utilizing regression analysis and various explanatory variables. It detailed the effectiveness of models like Linear Regression and Ridge Regularization, emphasizing their significance in employing machine learning techniques for addressing house price prediction challenges.

On a different note, Gergo Pinter et al. [14] introduced a pioneering machine learning approach that employed Call Detail Records (CDR) to investigate how mobility characteristics influence real estate price prediction. Utilizing AI and fundamental mobility entropy factors linked to dwellers' and workers' attributes, their model, employing a multi-layered perceptron (MLP) with particle swarm optimization (PSO), revealed substantial correlations between mobility factors and real estate prices. Similarly, Choujun Zhan et al. [15] highlighted the crucial role of accurate house price prediction models in shaping national real estate policies. Their research introduced a framework leveraging Hybrid Bayesian Optimization (HBO) models, including Stacking (HBOS), Bagging (HBOB), and Transformer (HBOT), aiming to enhance forecasting performance through hyperparameter tuning. A comparative analysis against benchmark models underscored the superior predictive performance of HBOS models, supported by a comprehensive multi-

source dataset encompassing Hong Kong real estate transactions from 1996 to 2021.

Despite the notable predictive capabilities of classical models utilized in previous research, attempts to enhance their efficiency have fallen short. In contrast, ML models have demonstrated varying degrees of effectiveness, particularly in real estate valuation and prediction within the residential domain, relying predominantly on readily available ML methodologies. However, existing literature has primarily focused on evaluating value and forecasting real estate using foundational ML techniques, overlooking the significance of integrating hybrid machine learning models derived from general-purpose models to streamline and elevate work practices. Presently, predictive models using ML for construction cost estimation in Thailand are gaining attention, albeit predominantly relying on widely used basic learning techniques. The implementation of hybrid learning methods lacks comprehensive comparative demonstrations, a crucial aspect within the learning domain. This knowledge gap necessitates the refinement of ML models for precise real estate price forecasting. Current studies endeavor to bridge these gaps comprehensively and address these challenges.

The aforementioned issues arise when business owners, project managers, or homeowners seek precise and detailed construction cost estimations, which consume considerable time and often require the involvement of engineers, architects, or estimators. Developing predictive models for flat-house construction cost estimation using cutting-edge ML techniques proves beneficial. This AI technique facilitates automated computer learning, enabling direct cost predictions from household area data, reducing the necessity for human intervention in the design process. Simultaneously, it maintains cost estimates closely aligned with traditional estimation methods

This research aims to achieve the objective of proposing innovative directions for low-rise building construction cost estimation using hybrid machine mearning (HML) techniques. The study endeavors to comprehensively compare the performance among various HML. Upon attaining these objectives, the study intends to present valuable insights into residential price estimations, aiming to enhance the accuracy and efficiency of real estate forecasts within the housing market ultimately. These advancements provide invaluable insights and tools for scholars, practitioners, engineers, architects, and investors aiming to gain a deeper comprehension of housing market dynamics. This deeper understanding enables more accurate predictions of market trends.

The remaining sections of this paper adhere to a structured framework. Section 2 provides an overview of the dataset and hybrid machine learning methodologies. Section 3 presents the empirical findings, culminating in Section 4, the conclusion.

2. METHODS

In this section, the dataset description for house price forecasting is defined. Following this, a data pre-processing method is proposed for house price prediction utilizing individual machine learning models and hybrid machine learning techniques. Within this framework, the analysis focuses on assessing the performance enhancement of 10 hybrid models for house price forecasting, evaluated using 8 measures to gauge predictive accuracy.

2.1. Dataset

The proposed model was applied using a dataset obtained from the Bureau of Public Works in Bangkok, focusing on low-rise housing and their corresponding prices. This dataset includes 15 key features defining different aspects of the low-rise buildings: construction cost (y), number of stories (x1), total usable area covering all floors (x2), area dedicated to bedrooms (x3), allocated bathroom area (x4), space for living rooms or restrooms (x5), kitchen and dining area (x6), laundry facilities

area (x7), balcony area (x8), space for stairs and corridors (x9), areas covered by roofing (x10, x11), parking area (x12), overall building height (x13), roof height (x14), and average floor height (x15). These features collectively describe various physical and functional attributes of the low-rise structures, forming the basis for the predictive model's analysis.

2.2. Data analysis

In this study, we conducted a comparative analysis of various HML algorithms like ANN, SVM, MLR, DT, and RF to predict construction costs for low-rise building projects. The aim was to identify the most precise and dependable method for estimating construction expenses. The dataset used in this analysis includes information on low-rise housing and their corresponding prices obtained from the Bureau of Public Works in Bangkok. It comprises 120 samples and encompasses 15 distinct attributes. Statistical measures of the dataset, including minimum, mean, maximum, and standard deviation, are visually represented in Figure 2.

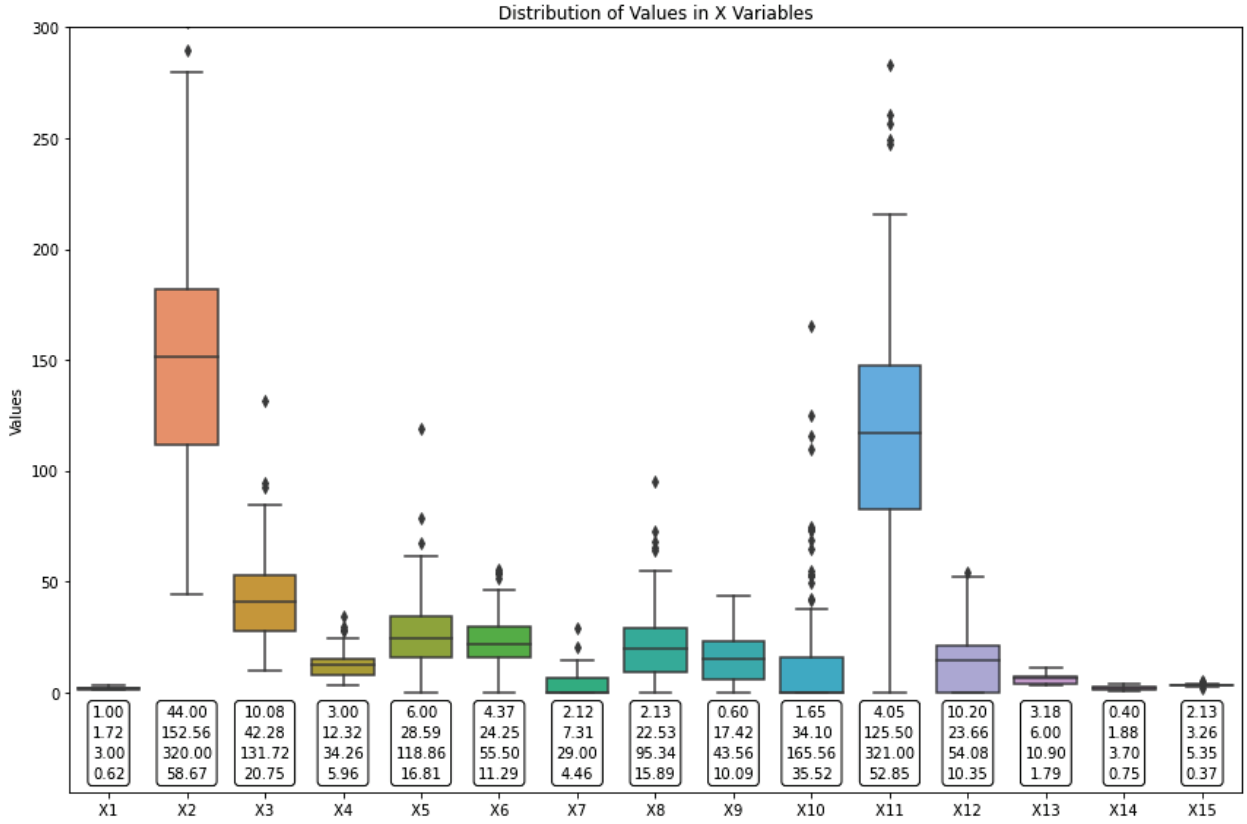

Figure 2. Visualization of the dataset

2.3. Data pre-processing

The meticulous data preprocessing steps involved managing missing data and precise segmentation into training and testing sets. The initial focus in predicting low-rise building prices was on selecting high-quality datasets sourced meticulously from reliable sources. These datasets encompass features like room area, building height, and amenities, all influential in determining the prices of low-rise buildings. Ultimately, the dataset was split into training, validation, and testing sets, facilitating model evaluation, training, hyperparameter fine-tuning, and accurate performance assessment.

2.4. Machine learning techniques base model

ML algorithms excel in modeling intricate and ambiguous systems, even when nonlinear relationships are unknown. In this study, five distinguished ML algorithms, namely ANNs,

SVMs, MLR, DT, and RF, are employed to create the proposed models. Leveraging these diverse algorithms enhances the predictive power of the approach. Additionally, detailed explanations of the design parameters for each algorithm are provided. The workflow process of the fundamental base ML is depicted in Figure 3.

Artificial Neural Networks (ANN) are a class of machine learning algorithms inspired by the human brain's neural structure. Comprising interconnected nodes (neurons) arranged in layers, ANNs are capable of learning complex patterns and relationships in data. They operate by receiving input data, processing it through these interconnected layers, and producing an output [14, 16].

Support Vector Machines (SVM) are supervised learning models used for classification and regression tasks. SVMs classify data by finding the optimal hyperplane that best separates different classes in a dataset [17]. They work well in both linearly and non-linearly separable datasets by transforming input data into higher dimensions to find an optimal boundary.

Multiple Linear Regression (MLR) is a statistical method used for modeling the relationship between a dependent variable and multiple independent variables [18]. It extends simple linear regression by accommodating more than one predictor variable, predicting a continuous output by estimating the coefficients of each predictor variable.

Decision Trees (DT) are tree-like structures used for both classification and regression tasks. They make decisions by splitting the dataset based on feature values, creating a tree structure of if-else decision nodes that lead to leaf nodes providing predictions [19].

Random Forest (RF) is an ensemble learning technique based on the concept of constructing multiple decision trees. It builds multiple trees and merges their predictions to improve accuracy and prevent overfitting by aggregating the predictions from various trees. Random forests are versatile and suitable for classification and regression problems [20].

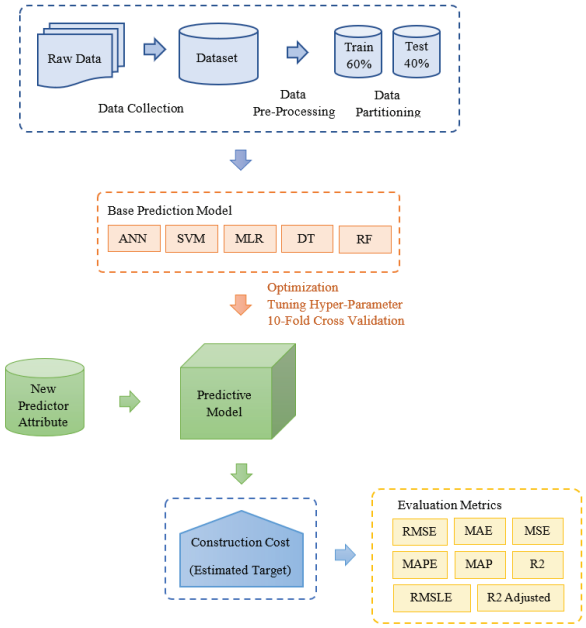

Figure 3. The classic base ML work flow

2.5. Hybrid machine learning model

Hybrid machine learning (HML) represents an innovative technique in machine learning where multiple models are amalgamated to enhance predictive accuracy, robustness, and generalization compared to employing a single model. Each hybrid method incorporates a distinct strategy to combine predictions from individual base models, employing various techniques. The workflow of HML is illustrated in Figure 4.

In the realm of predictive modeling, the quest for accuracy and robustness drives innovation. Recent advancements have shown that combining distinct machine learning algorithms into a hybrid model can significantly improve predictive performance. This study delves into the synthesis of hybrid models, each integrating two distinguished algorithms - an approach that has exhibited promising results in enhancing predictive power across various domains. Utilize a diverse set of ten hybrid models, including combinations like ANN-SVM, ANN-MLR, ANN-DT, ANN-RF, SVM-MLR, SVM-DT, SVM-RF, MLR-DT, MLR-RF, and DT-RF.

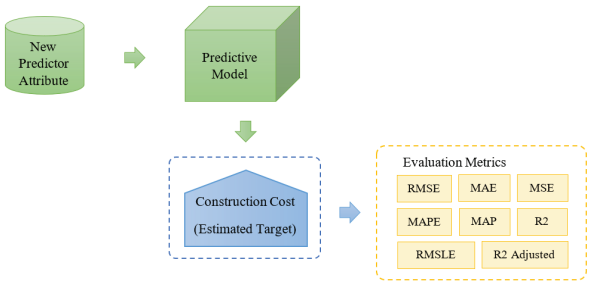

Figure 4. The HML work flow

ANN-SVM, the fusion of ANN and SVM aims to leverage the strength of both models. ANN, known for its ability to learn complex relationships in data, is combined with SVM, revered for its robustness in handling non-linear data. By merging ANN's capacity to comprehend intricate patterns with SVM's proficiency in handling high-dimensional datasets, the hybrid model offers a formidable approach to prediction tasks, effectively capturing nuanced relationships while maintaining resilience against overfitting.

ANN-RF, the hybridization of RF with ANN capitalizes on the strengths of both models. RF, a robust ensemble technique, brings its prowess in handling high-dimensional data and mitigating overfitting. By combining it with ANN's proficiency in learning complex relationships, this hybrid model endeavors to achieve improved generalization and predictive accuracy. The integration aims to refine predictions by mitigating biases from individual models, thus enhancing overall predictive performance.

ANN-DT, the fusion of ANN and DT seeks to leverage ANN's capacity for learning complex relationships and DT's interpretability and handling of non-linearities. By combining ANN's deep learning capabilities with DT's tree-based structure, this hybrid approach aims to capture intricate patterns while ensuring transparency and ease of interpretation in the decision-making process.

ANN-MLR, the amalgamation of ANN with MLR attempts to amalgamate ANN's ability to capture complex data patterns with MLR's simplicity in handling linear relationships. This union endeavors to exploit ANN's proficiency in learning intricate features while augmenting it with the interpretability and clarity of linear relationships provided by MLR, creating a balanced predictive model.

SVM-MLR, the coupling of SVM and MLR strives to unite SVM's robustness in handling complex data structures and MLR's capacity for understanding linear relationships. By merging the strengths of SVM's effective classification boundaries with MLR's straightforward interpretation, this hybrid approach aims to create a model that adeptly navigates both linear and non-linear aspects in data.

SVM-DT, the fusion of SVM and DT capitalizes on SVM's ability to handle high-dimensional data and DT's interpretability in handling complex decision boundaries. This pairing aims to amalgamate SVM's resilience against overfitting with DT's capacity to partition data into smaller subsets, providing a model capable of handling complex data structures while retaining interpretability.

SVM-RF, the fusion of SVM with RF aims to merge SVM's ability to handle high-dimensional data and RF's ensemble-based learning. By combining SVM's robustness in creating effective decision boundaries with RF's ensemble of diverse decision trees, this hybrid model aspires to enhance predictive performance by leveraging the strengths of both algorithms.

MLR-DT, a fundamental tool for understanding linear relationships, is coupled with DT, famed for its interpretability and non-parametric nature. The combination aims to harness the interpretive simplicity of linear regression while augmenting it with DT's capability to handle non-linearities and interactions within the data. This amalgamation promises a balanced model that navigates both linear and non-linear aspects, providing a comprehensive view of the predictive landscape.

MLR-RF, the fusion of MLR with RF seeks to unify MLR's ability to model linear relationships with RF's ensemble-based learning approach. MLR, known for its simplicity and interpretability in modeling linear relationships,

is combined with RF's ensemble of decision trees, which excels in capturing complex nonlinearities and interactions within data. This amalgamation aims to create a hybrid model that harmonizes the interpretability of linear regression with the predictive power of ensemble-based learning.

DT-RF, the fusion of DT with RF seeks to combine DT's tree-based structure and interpretability with RF's ensemble learning approach. This hybridization aims to leverage the interpretability and ease of understanding provided by DT while harnessing the robustness and predictive power of RF's ensemble learning, potentially enhancing overall model performance.

In conclusion, the amalgamation of these leading machine learning algorithms into hybrid models presents a compelling approach to enhancing predictive capabilities. The combinations offer complementary strengths, effectively addressing various complexities inherent in diverse datasets. The pursuit of hybrid models, combining two powerful algorithms at a time, marks a significant stride toward achieving more accurate and resilient predictive models in the architecture of machine learning.

2.6. Performance measure

Mean Absolute Error (MAE) is the average of the absolute differences between the predicted and actual values. It measures the average magnitude of errors without considering their direction [19].

Mean Squared Error (MSE) is a measure of the average squared difference between the actual and predicted values [19]. It squares the difference between predicted and actual values for each data point, sums those squares, and then divides by the number of data points (or samples). It penalizes larger errors heavily due to the squaring effect.

Root Mean Squared Error (RMSE) is the square root of the MSE. It provides an interpretable metric in the same units as the target variable [13]. RMSE is more sensitive to outliers compared to MAE.

R-squared (R2) represents the proportion of variance in the dependent variable that is predictable from the independent variables in the model. It ranges from 0 to 1, with higher values indicating a better fit of the model to the data [18, 21].

Mean Absolute Deviation (MAD) calculates the mean of the absolute deviations between predicted and actual values. It measures the average difference between predicted and actual values.

Mean Absolute Percentage Error (MAPE) computes the average percentage difference between predicted and actual values. It's useful for understanding the relative error between predictions and true values [19].

Root Mean Squared Logarithmic Error (RMSLE) calculates the average difference of the logarithmic values of predicted and actual values. It's commonly used when the target variable has a wide range of values [20].

Adjusted R-squared (R2 Adjusted). Adjusted is a modified version of R2 that adjusts for the number of predictors in the model. It penalizes the addition of unnecessary predictors that do not improve the model significantly.

The predictive performance of the model is assessed using eight evaluation measures, defined as follows

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_{i(pred)} - y_{i(actual)}\right| \quad (1)$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_{i(pred)} - y_{i(actual)}\right)^2 \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_{i(pred)} - y_{i(actual)}\right)^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum (y_{i(pred)} - y_{i(actual)})^2}{\sum (y_{i(\text{average})} - y_{i(actual)})^2} \quad (4)$$

$$MAD = \frac{1}{N}\sum_{i=1}^{N} median_n\left(y_{i(actual)} - y_{i(pred)}\right) \quad (5)$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|y_{i(actual)} - y_{i(pred)}\right|}{\max(\tau, \left|y_{i(actual)}\right|)} \quad (6)$$

$$RMSLE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log_e(1+y_{i(actual)}) - \log_e(1+y_{i(\text{pred})}))^2} \quad (7)$$

$$R^2_{adjusted} = 1 - \left[\frac{(1-R^2)(n-1)}{n-m-1}\right] \quad (8)$$

where yi(actual) is the actual low-rise building construction cost, yi(pred) is the predicted low-rise building construction cost, yi(average) is the average low-rise building construction cost, N is the total number of observations, m is the number of sample features and τ is an arbitrarily small but strictly positive number to avoid undefined results [15, 18].

The success of predicting construction cost estimation relies significantly on selecting the most fitting algorithm. The evaluation of various algorithms is carried out meticulously, employing pertinent assessment metrics like R2, MSE, or MAE on the validation dataset. The algorithm that displays exceptional predictive capabilities and robust generalization becomes the chosen prediction model. Overall, these methodologies span crucial phases from data selection and preprocessing to model training and algorithm selection, providing a comprehensive solution for low-rise building price prediction. This systematic approach enables real estate professionals, investors, and analysts to make well-informed decisions backed by accurate predictions of low-rise building prices.

3. RESULT AND DISCUSSION

In this section, we delve deeply into the outcomes derived from the meticulous development and training of machine learning hybrid models, utilizing a pre-processed dataset. The focal point lies in presenting a comprehensive comparative analysis of the performance of all ten hybrid machine learning algorithms. This comprehensive analysis encapsulates an all-encompassing parameter tuning process aimed at bolstering the robustness in performance evaluation. Furthermore, employing 10-fold cross-validation enhances the reliability of this performance assessment.

The primary objective of this study is to forecast low-rise building construction costs utilizing the advantages of ten sets of hybrid machine learning models. To measure their efficacy, we employed eight distinct statistical metrics, as detailed in Table 1, aiming to assess the accuracy of each model in estimating the housing construction costs in Thailand. As indicated in Table 1, the ANN-DT model emerged as the most prominent hybrid model, exhibiting an impressive accuracy of 0.921. Following closely, the ANN-MLR showcased a commendable accuracy of 0.916, while both the ANN-RT and MLR-DT demonstrated accuracies of 0.907. Conversely, the models utilizing the SVM algorithms displayed relatively lower accuracies, especially the SVM-RF model, registering the lowest accuracy at 0.696.

Table 1. The summary of the HML algorithms performance

No.	Model	R2	R2_Adj	MSE	RMSE	MAE	MAD	MAPE	RMSLE
1	ANN-SVM	0.804	0.713	85611185529	292593.89	195276.88	195276.88	0.18	0.24
2	ANN-MLR	0.916	0.877	36642276522	191421.72	135165.07	135165.07	0.10	0.13
3	**ANN-DT**	**0.921**	**0.885**	**34164053244**	**184835.21**	**125468.65**	**125468.65**	**0.11**	**0.15**
4	ANN-RF	0.905	0.861	41433259536	203551.61	137532.68	137532.68	0.11	0.15
5	SVM-MLR	0.753	0.637	108144129649	328852.75	198573.01	198573.01	0.16	0.21
6	SVM-DT	0.759	0.646	105538683718	324867.18	192032.70	192032.70	0.18	0.24
7	SVM-RF	0.696	0.554	132884310435	364533.00	211720.14	211720.14	0.18	0.25
8	MLR-DT	0.907	0.864	40622706349	201550.75	126171.77	126171.77	0.09	0.12
9	MLR-RF	0.875	0.816	54831353216	234160.96	143613.72	143613.72	0.09	0.12
10	DT-RF	0.868	0.807	57601528945	240003.19	138988.05	138988.05	0.11	0.16

Consequently, these research findings enable us to draw two significant conclusions. Firstly, the ANN-DT model outperforms all other models, consistently securing the top rank across almost all eight-evaluation metrics. This indicates that ANN-DT demonstrates increased stability and superior overall performance within the housing price dataset. Secondly, these findings suggest that the ANN-DT hybrid model is the most suitable for estimating low-rise building construction costs in Thailand, surpassing other hybrid learning models.

The comparison of accuracy performance among the 10 hybrid machine learning models is depicted in Figure 5. This representation illustrates their effectiveness in predicting the construction cost of low-rise residential buildings, showcasing varying predictive capabilities. However, aiming for predictions with an error rate of less than 10 percent, aligned with AACE's recommendations for finished design and model approximation, reveals only 4 models that meet this criterion. These models consist of ANN-DT, ANN-MLR, ANN-RF, and MLR-DT.

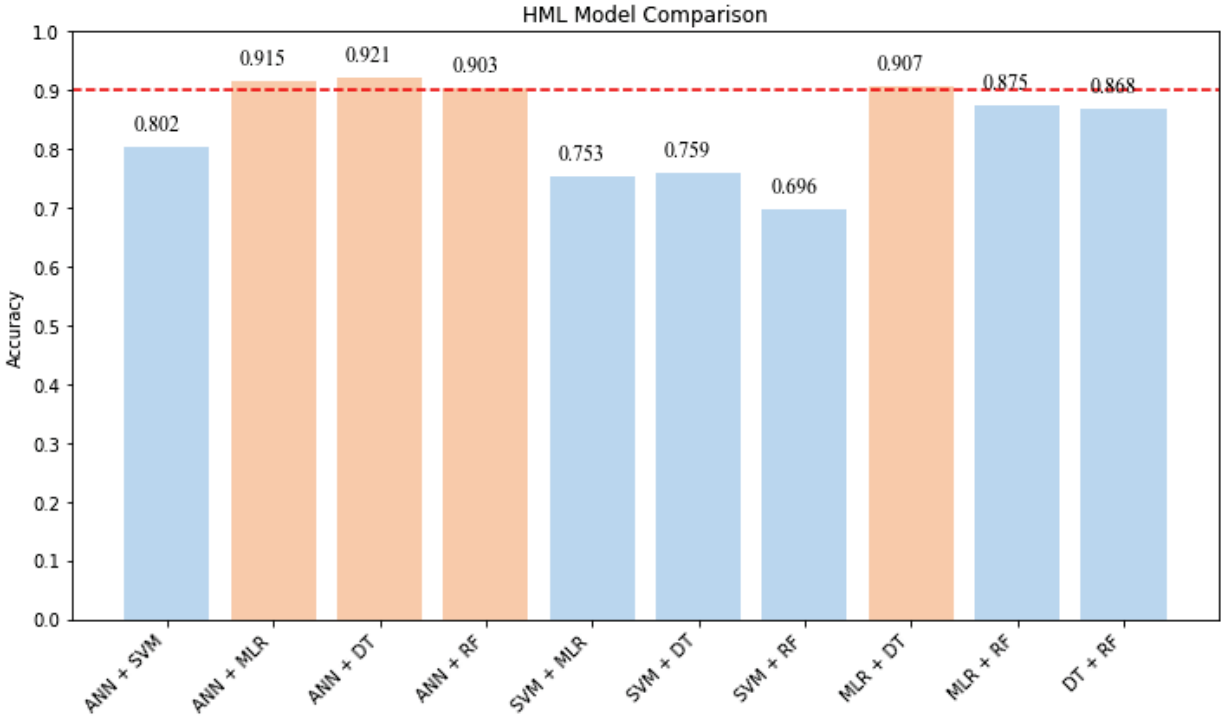

Figure 5. Comparison of accuracy among the HML models

In summary, the research involved evaluating predictions made by 10 hybrid machine learning models employing diverse methodologies to forecast the prices of low-rise buildings. The findings emphasize the exceptional performance of ANN-DT, positioning it as an appealing choice for practical applications in this domain. Furthermore, models such as ANN-MLR, ANN-RF, and MLR-DT demonstrate outstanding accuracy, surpassing acceptable standards for estimating construction costs follow the AACE recommended. This substantiates their standing as valuable options worthy of consideration.

4. CONCLUSION

In conclusion, this study harnessed the predictive potential of ten HML models to anticipate low-rise building construction cost. Predicting construction costs for residential housing holds significant importance within the

economic landscape. Extensive research has explored machine learning applications in modeling house price predictions. However, this study encounters specific limitations. Firstly, diverse machine learning models exhibit distinct strengths and weaknesses, and a model's performance in a specific metric or dataset doesn't ensure universal efficacy. Secondly, enhancing machine learning model performance heavily relies on hyperparameters, posing challenges in optimizing models through traditional grid or random search methods.

To address prevailing challenges in current house price prediction methodologies, characterized by high similarity in price trends across different regions, this study introduces a comprehensive and innovative framework for accurate house price forecasting. Initially, basic models including ANN, SVM, MLR, DT, and RF were utilized to develop ten hybrid machine learning models incorporating techniques such as ANN-SVM, ANN-MLR, ANN-DT, ANN-RF, SVM-MLR, SVM-DT, SVM-RF, MLR-DT, and DT-RF. Next, eight evaluation metrics including MSE, RMSE, MAE, $R2$, MAD, MAPE, $R2adjusted$, RMSLE were employed to analyze and compare the performance of these models.

The findings pinpointed the remarkable accuracy of ANN-DT with a high score of 0.921 among all tested models. This hybrid technique amalgamated expertise from different models, showcasing outstanding results. Additionally, ANN-MLR, ANN-DT, and MLR-DT demonstrated competitive accuracy, scoring 0.916, 0.907, and 0.904, respectively. These outcomes underscore significant performance enhancements achievable through mass techniques compared to details designing and tweaking models to limit deviations to under 10%, as recommended by AACE.

The study contributes valuable insights applicable to both academia and industry, particularly in refining house price prediction accuracy. It represents a significant stride in real estate market research by addressing existing limitations and establishing a foundational methodological framework for future house price prediction studies. Although this study demands substantial computational resources, future endeavors aim to explore more efficient model architectures, optimization techniques, and increased data utilization. Further research will encompass three facets: integrating a multi-source data fusion framework, employing diverse datasets for comprehensive model performance assessment, and exploring factors influencing real estate values to provide policy recommendations.

ACKNOWLEDGEMENTS

This research project was financially supported by Mahasarakham University of Thailand under Grant No. 6603024.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. **M.-Y. Cheng, H.-C. Tsai, and W.-S. Hsieh**, "Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model," Automation in Construction, vol. 18, no. 2, pp. 164-172, 2009.
2. **S.S. Arage and N.V. Dharwadkar,** "Cost estimation of civil construction projects using machine learning paradigm," in 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2017: IEEE, pp. 594-599.
3. **Sirithikon Sittikarnkul,** "Construction Cost Estimation for Government Building Using Prediction Modeling Techniques," Master of Engineering, Graduate School, Chiang Mai University, 2021.

4. **Kawee Wangnivejankul**, Construction Cost Estimating. Bangkok: SE-EDUCATION, 2023.
5. **R.P. Huehmer**, "Detailed estimation of desalination system cost using computerized cost projection tools," in 12th annual conference, DesalinationVisions for the Future, 2011, pp. 14-15.
6. **R.P.N.R.-.** AACE International, Cost Estimate Classification System - As Applied for the Petroleum Exploration and Production Industry. Morgantown, WV: AACE International, August 7, 2020.
7. **P. Chandu and N.B. Devi,** "Improved Prediction Accuracy of House Price Using Decision Tree Algorithm over Linear Regression Algorithm," in 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), 2023: IEEE, pp. 1-6.
8. **Y. Xu, Y. Zhou, P. Sekula, and L. Ding,** "Machine learning in construction: From shallow to deep learning," Developments in the built environment, vol. 6, p. 100045, 2021.
9. **L. Kittisak and W. Kittipol,** "The Prediction of Low-Rise Building Construction Cost Estimation Using Extreme Learning Machine," Adv. technol. innov., Dec. 2023.
10. **S. Tayefeh Hashemi, O. M. Ebadati, and H. Kaur,** "Cost estimation and prediction in construction projects: A systematic review on machine learning techniques," SN Applied Sciences, vol. 2, pp. 1-27, 2020.
11. **A.O. Elfaki, S. Alatawi, and E. Abushandi**, "Using intelligent techniques in construction project cost estimation: 10-year survey," Advances in Civil engineering, vol. 2014, 2014.
12. **S. Lu, Z. Li, Z. Qin, X. Yang, and R.S.M. Goh**, "A hybrid regression technique for house prices prediction," in 2017 IEEE international conference on industrial engineering and engineering management (IEEM), 2017: IEEE, pp. 319-323.
13. **S. Chiramel, D. Logofătu, J. Rawat, and C. Andersson,** "Efficient Approaches for House Pricing Prediction by Using Hybrid Machine Learning Algorithms," in Intelligent Information and Database Systems: 12th Asian Conference, ACIIDS 2020, Phuket, Thailand, March 23–26, 2020, Proceedings 12, 2020: Springer, pp. 85-94.
14. **G. Pinter, A. Mosavi, and I. Felde,** "Artificial intelligence for modeling real estate price using call detail records and hybrid machine learning approach," Entropy, vol. 22, no. 12, p. 1421, 2020.
15. **C. Zhan, Y. Liu, Z. Wu, M. Zhao, and T. W. Chow**, "A hybrid machine learning framework for forecasting house price," Expert Systems with Applications, vol. 233, p. 120981, 2023.
16. **J. Kalliola, J. Kapočiūtė-Dzikienė, and R. Damaševičius,** "Neural network hyperparameter optimization for prediction of real estate prices in Helsinki," PeerJ computer science, vol. 7, p. e444, 2021.
17. **B.B. Nair, V. Mohandas, and N. Sakthivel,** "A genetic algorithm optimized decision tree-SVM based stock market trend prediction system," International journal on computer science and engineering, vol. 2, no. 9, pp. 2981-2988, 2010.
18. **K. Lathong and K. Wisaeng,** "The Prediction of Low-Rise Building Construction Cost Estimation Using Extreme Learning Machine," Adv. technol. innov., Dec. 2023.
19. **S. Papadopoulos, E. Azar, W.-L. Woon, and C.E. Kontokosta,** "Evaluation of tree-based ensemble learning algorithms for building energy performance estimation," Journal of Building Performance Simulation, vol. 11, no. 3, pp. 322-332, 2018.
20. **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.

21. **L. El Mouna, H. Silkan, Y. Haynf, M.F. Nann, and S.C. Tekouabou**, "A Comparative Study of Urban House Price Prediction using Machine Learning Algorithms," in E3S Web of Conferences, 2023, vol. 418: EDP Sciences, p. 03001.

СПИСОК ЛИТЕРАТУРЫ

1. **M.-Y. Cheng, H.-C. Tsai, and W.-S. Hsieh**, "Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model," Automation in Construction, vol. 18, no. 2, pp. 164-172, 2009.
2. **S.S. Arage and N.V. Dharwadkar,** "Cost estimation of civil construction projects using machine learning paradigm," in 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2017: IEEE, pp. 594-599.
3. **Sirithikon Sittikarnkul,** "Construction Cost Estimation for Government Building Using Prediction Modeling Techniques," Master of Engineering, Graduate School, Chiang Mai University, 2021.
4. **Kawee Wangnivejankul**, Construction Cost Estimating. Bangkok: SE-EDUCATION, 2023.
5. **R.P. Huehmer**, "Detailed estimation of desalination system cost using computerized cost projection tools," in 12th annual conference, DesalinationVisions for the Future, 2011, pp. 14-15.
6. **R.P.N.R.-.** AACE International, Cost Estimate Classification System - As Applied for the Petroleum Exploration and Production Industry. Morgantown, WV: AACE International, August 7, 2020.
7. **P. Chandu and N.B. Devi,** "Improved Prediction Accuracy of House Price Using Decision Tree Algorithm over Linear Regression Algorithm," in 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), 2023: IEEE, pp. 1-6.
8. **Y. Xu, Y. Zhou, P. Sekula, and L. Ding,** "Machine learning in construction: From shallow to deep learning," Developments in the built environment, vol. 6, p. 100045, 2021.
9. **L. Kittisak and W. Kittipol,** "The Prediction of Low-Rise Building Construction Cost Estimation Using Extreme Learning Machine," Adv. technol. innov., Dec. 2023.
10. **S. Tayefeh Hashemi, O. M. Ebadati, and H. Kaur,** "Cost estimation and prediction in construction projects: A systematic review on machine learning techniques," SN Applied Sciences, vol. 2, pp. 1-27, 2020.
11. **A.O. Elfaki, S. Alatawi, and E. Abushandi**, "Using intelligent techniques in construction project cost estimation: 10-year survey," Advances in Civil engineering, vol. 2014, 2014.
12. **S. Lu, Z. Li, Z. Qin, X. Yang, and R.S.M. Goh**, "A hybrid regression technique for house prices prediction," in 2017 IEEE international conference on industrial engineering and engineering management (IEEM), 2017: IEEE, pp. 319-323.
13. **S. Chiramel, D. Logofătu, J. Rawat, and C. Andersson,** "Efficient Approaches for House Pricing Prediction by Using Hybrid Machine Learning Algorithms," in Intelligent Information and Database Systems: 12th Asian Conference, ACIIDS 2020, Phuket, Thailand, March 23–26, 2020, Proceedings 12, 2020: Springer, pp. 85-94.
14. **G. Pinter, A. Mosavi, and I. Felde,** "Artificial intelligence for modeling real estate price using call detail records and hybrid machine learning approach," Entropy, vol. 22, no. 12, p. 1421, 2020.
15. **C. Zhan, Y. Liu, Z. Wu, M. Zhao, and T. W. Chow**, "A hybrid machine learning framework for forecasting house price," Expert Systems with Applications, vol. 233, p. 120981, 2023.
16. **J. Kalliola, J. Kapočiūtė-Dzikienė, and R. Damaševičius,** "Neural network hyperparameter optimization for prediction

of real estate prices in Helsinki," PeerJ computer science, vol. 7, p. e444, 2021.
17. **B.B. Nair, V. Mohandas, and N. Sakthivel,** "A genetic algorithm optimized decision tree-SVM based stock market trend prediction system," International journal on computer science and engineering, vol. 2, no. 9, pp. 2981-2988, 2010.
18. **K. Lathong and K. Wisaeng,** "The Prediction of Low-Rise Building Construction Cost Estimation Using Extreme Learning Machine," Adv. technol. innov., Dec. 2023.
19. **S. Papadopoulos, E. Azar, W.-L. Woon, and C.E. Kontokosta,** "Evaluation of tree-based ensemble learning algorithms for building energy performance estimation," Journal of Building Performance Simulation, vol. 11, no. 3, pp. 322-332, 2018.
20. **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.
21. **L. El Mouna, H. Silkan, Y. Haynf, M.F. Nann, and S.C. Tekouabou**, "A Comparative Study of Urban House Price Prediction using Machine Learning Algorithms," in E3S Web of Conferences, 2023, vol. 418: EDP Sciences, p. 03001.

Kittipol Wisaeng received a Ph.D. in Electrical and Computer Engineering from Thailand in 2016. He is currently an Associate Professor in Computer Science. His main research interests center on advancing artificial intelligence methods and systems integration in applications that lie predominantly in medical image segmentation. He has carried out a unique portfolio of research programs, generally focusing on deep learning and machine learning. He addresses such issues as optimization techniques, retinal segmentation, breast cancer detection, blood vessel extraction, exudates identification, optic disc localization, and brain tumor classification in challenging environments. He has published over 26 technical papers in this area and has seven H-index. He received the senior researcher award, outstanding publication award, the highest publication of research in the Scopus database, the highest h-index award, and the highest citation per article organized by Mahasarakham University. He was a Guest Reviewer of Applied Soft Computing in 2014, IEEE Transactions on Instrumentation and Measurement in 2023, Journal of Intelligent & Fuzzy Systems in 2022, Soft Computing in 2020, Computers, Materials & Continua in 2022, Intelligent Automation & Soft Computing in 2023, and IET Image Processing in 2023. He can be contacted at email: kittipol.w@acc.msu.ac.th

Kittisak Lathong rereceived a B.Eng. degree in Civil Engineering from Srinakharinwirot University, Thailand, a B.P.H. degree in Occupational Health and Safety from Sukhothai Thammathirat Open University, Thailand, a M.Eng. degree in Civil Engineering from KMITL, Thailand, a M.P.A. degree in Public Administration from Ramkhamhaeng University, Thailand and culminating with a Ph.D. degree in Business Administration and Innovation from The Mahasarakham University, Thailand. With over 15 years of dedicated experience in the fields of design and construction, he possesses exceptional expertise. His professional proficiency is a testament to his extensive hands-on involvement and comprehensive understanding of the industry. His research pursuits encompass various topics, notably civil engineering, construction engineering, building information modeling, machine learning, and artificial intelligence. His dedication to advancing these domains underscores his commitment to driving innovation and progress within the field. He can be contacted at email: kittisak.lathong@gmail.com.

Киттипол Висаенг, доктор философии в области электротехники и вычислительной техники, доцент в области компьютерных наук. Электронная почта: kittipol.w@acc.msu.ac.th.

Киттисак Латхонг, магистр в области гражданского строительства, доктор философии в области управления бизнесом и инноваций, Электронная почта: kittisak.lathong@gmail.com.