

House Type Specification for Housing Development Project Using Machine Learning Techniques: A Study From Bangkok Metropolitan Region, Thailand

Kongkoon Tochaiwat*, Patcharida Seniwong

Faculty of Architecture and Planning, Thammasat University, Thailand

*Corresponding e-mail: kongkoon@tu.ac.th

Received 2023-03-06; Revised 2024-02-07; Accepted 2024-02-09

ABSTRACT

Specifying the house type of a housing development project is extremely necessary. However, the determination of a project type nowadays has become a delicate matter, requiring the expertise and knowledge of seasoned project developers. This study aimed to apply four machine learning techniques: Decision Tree, Random Forest, Gradient Boosted Tree and Ensemble Classifier, to analyze the data from 179 housing estate projects collected from market reports of real estate companies in Thailand, with a focus on selecting projects with average monthly sales rates that are higher than the average of all acquired projects. This process resulted in a reduced dataset of 59 projects, including 31 townhouses, 22 single-family houses, and six semi-detached houses. As a result, the Ensemble Classifier model has the highest accuracy of 90.91%. The factors most influential in identifying the type of project are the distances from a main road, sky train station, bus station, hospital, and department store. Single-detached house projects are suitable for locations with high potential. The ideal location should be in proximity to a main road, bus station, department store, and hospital. In addition, townhouse projects are ideal for medium-potential locations that are not near shopping malls, but still require proximity to a hospital, sky train station, or bus station. Ultimately, semi-detached house projects are ideal for medium-potential locations that require proximity to a main road for convenient access to sky train station or public transportation, depending on the specific context.

Keywords: prediction model, house type, housing development project, machine learning, ensemble classifier

INTRODUCTION

Housing is considered a fundamental aspect of human life and is recognized as one of the four essential components of living, in accordance with the cultural beliefs of Thai people. Presently, the residential development sector continues to exhibit a significant rate of growth (Real Estate Information Center [REIC], 2018). Real estate entrepreneurs commonly prioritize housing project development in their business. The precise selection of housing type based on informed criteria is crucial to ensure successful sales outcomes for launched projects (Tochaiwat, 2020).

The process of selecting the housing types by real estate entrepreneurs is primarily dependent on the knowledge, experience, and expertise of relevant specialists (Kamglin & Tochaiwat, 2014; Real Asset, 2021). However, the availability of experienced and knowledgeable specialists is limited, as expertise in this field requires a significant accumulation of experience and can still be prone to error. Physical factors, particularly land conditions, change from project to project in the housing development industry. To address this issue, some individuals are developing knowledge and models to support decision-making in these projects (Baldominos et al., 2018; Shen & Ross, 2021). For example, the research of Habib and Kockelman (2008) presented the issue of setting house styles in relation to the habitable location of the population in each area. In addition, Phe and Wakely (2000) explained that housing style quality and development in varying contexts can impact urban development and housing project success.

A review of the relevant literature revealed that the existing model for determining housing types in real estate projects was not adequate when considering physical factors, such as location and other properties of the project. Therefore, the researchers' interest lies in how to apply Machine Learning Theory to determining project type, based on physical factor data collected from secondary sources that entrepreneurs can assess, in order to generate higher sales rates. In the housing development process, sales rate, i.e., the average number of units sold per month, is a very important collectable index because it reflects the monthly cash inflow to the project (Tochaiwat, 2020). Machine Learning is a

valuable technique for real estate entrepreneurs in analyzing complex and multi-related factors (Big Data Thailand, 2020). Such a technique, therefore, is suitable for the identification of physical factors for housing estate projects as it is able to help determine suitable project types, create knowledge and analytical tools, and respond to the demands of housing project buyers, as well as enhance the competitiveness and increase the chances of success for the business in the future.

LITERATURE REVIEW

Specification of a Housing Development Project

The specification of a housing development project is an essential step as it establishes the objectives that guide the subsequent phases of the workflow. Specifying housing types will aid real estate entrepreneurs in creating a project development policy that is more responsive to the needs of residents (Ching, 2023; Egemen, 2022; Henilane, 2016). For housing estate projects, entrepreneurs need to acquire and verify land that meets their plans before exactly specifying any physical characteristics, considering it a critical factor in the real estate development process (Terra Media and Consulting, 2017). Moreover, a project feasibility study should be conducted before a land purchase agreement is reached to ensure the ultimate financial benefits for both the entrepreneur and the inhabitants (Luque et al., 2019).

A feasibility study for a housing estate development project typically comprises three components: (1) a market feasibility study, (2) a physical project feasibility study, and (3) a financial feasibility study (Tochaiwat, 2020). The literature review indicated that market feasibility comprises an analysis of economic trends and market demand. The identification of the target customer group in a housing estate project facilitates a deeper understanding of customer needs and informs the physical feasibility study; this study assesses the physical constraints of the land and environment and considers both the desired house style and the lifestyle of the

intended residents. Financial feasibility studies play a crucial role for real estate entrepreneurs in ensuring the sustainability of their business operations (Rissanen et al., 2020). However, amidst a competitive landscape, real estate entrepreneurs must conduct a feasibility analysis of housing estate projects to accurately determine patterns that meet the needs of their target customer groups (Tochaiwat, 2018). The aforementioned points align with the findings of Mariadas et al. (2019), who discovered that conducting feasibility studies and acquiring land are crucial processes in the development of housing projects, particularly with regards to determining the house style and residential location.

In Thailand, the real estate sector contributed to 4.8% of Thailand's GDP in 2021 (sharply decreased from 8% in 2019, the pre-pandemic year). There is major demand for housing in the Bangkok Metropolitan Region. Figure 1 shows the boundaries of Bangkok and the five surrounding provinces, i.e., Samut Prakan, Pathum Thani, Nakhon Pathom, Samut Sakhon, and Nonthaburi. The ratios of the number of land subdivision permits in Bangkok Metropolitan Region to the whole country were higher than 50% during 2015 to 2019. The ratio dropped a little during COVID-19 crisis and became about 48% in 2021 (Klinchuanchun, 2022).

There are three types of houses generally developed in the Bangkok Metropolitan Region: single-detached houses, twin houses and townhouses, as shown in Figure 2.

Figure 1

Bangkok Metropolitan Region



Note. Adapted from *Bangkok Metropolitan Region Map*, by Google Map (www.google.co.th/maps). Copyright 2023 by Google LLC.

Figure 2

Types of Houses: (a) single-detached houses, (b) twin houses and (c) townhouses



Note. Type (a) from *Single-Detached Houses*, by Lalin Property Plc., 2022 (www.lalinproperty.com/properties/lanceo-crib-2-bangna-suvarnabhumi/). Copyright 2022 by Lalin Property Plc. Type (b) from *Twin Houses*, by DDProperty, 2020 (www.ddproperty.com/โครงการ/7201-britania-bangna-suvarnabhumi-บริหารเนียบางนา-สุวรรณภูมิ). Copyright 2020 by DDProperty. Type (c) from *Townhouses*, by DDProperty, 2017 (www.ddproperty.com/ข่าวอสังหาริมทรัพย์-บทความ/2017/7/156142/ภาครัฐส่งสัญญาณปลุกอสัง). Copyright 2017 by DDProperty.

Factors Affecting Patterns of Housing Development Projects

The essential characteristics of the factors utilized in determining the pattern of a housing project must reflect the long-term quality and viability of the project's housing. These factors must drive competitiveness and demonstrate the sustainability of the organization, leading to profitability for entrepreneurs and enhancing the benefits for residents (Budhathoki, 2015; Zhang et al., 2018). The literature review revealed that the suitability of the house type for residents is a key factor in building their confidence in their purchase, thereby serving as a guarantee of the house's quality (Moghayedi et al., 2021; Torbica & Stroh, 2000). This finding aligns with the research of Elsinga and Hoekstra (2005), which demonstrated that the type of project is a significant consideration in the buying process and is a major contributor to the residents' satisfaction. The characteristics of a well-designated project include the following details:

Project Factors

According to the literature review, the critical factors affecting the design of the house were found to be: (1) the type of property developer and (2) land appraisal prices (Kung, 2022). In

addition, Araloyin and Olatoye (2011) found that the reputation and credibility of entrepreneurs play a significant role in home buying decisions, as trust and buyers' experiences are critical for ensuring the quality of a high-value asset such as a home. This is consistent with the findings of Koklic and Vida (2009), who noted that brand and entrepreneur reputation have an impact on the residents' home buying choices in determining the project design to close a sale. Moreover, the land appraisal price plays a crucial role in the land selection process of entrepreneurs as it has economic and legal implications that impact project type determination and home purchase decisions of residents in the project area (DDproperty, 2020). According to Nakamura (2019), land appraisal value is a critical indicator of land worth that highlights the sustainability of environmental, economic, and social aspects in the area. This is a crucial factor in fostering the development of project models to meet the needs of the target resident group, as envisioned by entrepreneurs and organizations.

The literature review also revealed that land appraisal prices and entrepreneurs were found to be factors that correlate with future house pricing (Li & Wu, 2014; Tochaiwat, 2020). Therefore, these factors are crucial in determining the type of residential project, as they can indicate the quality of the living space and impact future residents' purchasing choices.

Location Factors

According to a literature review, location emerged as the most crucial factor impacting home prices and satisfying the needs of residents, often serving as the primary consideration in determining the type of housing (Luttik, 2000; Mulyano et al., 2020; Vandeviver & Bernasco, 2020; Nursoleh, 2022). Therefore, in determining the project model, entrepreneurs must consider the location factor, as it plays a crucial role in indicating the potential of the area and influencing future residents (Maoludyo & Aprianingsih, 2015). The researchers found that the criteria for selecting a location for the project design is dependent on the proximity to surrounding landmarks and compliance with legal requirements (Rymarzak & Siemińska, 2012). Additionally, entrepreneurs should consider the demographic characteristics of each area when determining the design of their project, as they can vary significantly between locations (Zeng, 2013).

Regarding the literature review, the proximity to places of employment or essential places (e.g. shopping malls, mass transit stations, and utilities) significantly influences housing projects in a specific location (Acheampong, 2018; Khan et al., 2017; Mete & Yomralioglu, 2022). In this study, therefore, two groups of factors were used to determine the characteristics of the locations: (1) the project location group, consisting of the distance from the project to key places and (2) the project zone, comprising ten zones in the Bangkok Metropolitan Region, namely: (1) Northern Bangkok, (2) Inner Bangkok, (3) Eastern Bangkok, (4) Thonburi, (5) Bangkok Central Business District (CBD), (6) Nakhon Pathom, (7) Nonthaburi, (8) Pathum Thani, (9) Samut Prakan, and (10) Samut Sakhon (REIC, 2022).

In conclusion, the literature review indicated that the specification of housing projects is influenced by four main factors: (1) Entrepreneurs, (2) Land appraisal values, (3) Project locations, and (4) Project zones.

Machine Learning Technique

The machine learning technique is a systematic process of analyzing data using various methods. The provision of quality data results in more accurate analysis, thereby positively impacting the performance of subsequent studies (Mahesh, 2020). In this study, a literature review was conducted to determine the most suitable model for the data set to achieve the objectives of the project design. It was found that four techniques were suitable in terms of the research objective: (1) Decision Tree, (2) Random Forest, (3) Gradient Boosted Tree, and (4) Ensemble Classification.

According to Gavankar and Sawarkar (2017), the Decision Tree model is effective in grouping data from complex problem analysis. This is supported by Patel and Rana (2014) who asserted that the Decision Tree model is suitable for identifying long-term trends. Additionally, Belgiu and Drăguț (2016) and Speiser et al. (2019) showed the efficacy of the Random Forest model in accurately segmenting multiple data sets and providing flexibility in data analysis. This is in accordance with the findings of Yang (2018), who reported that the Random Forest model demonstrates good scalability and processing capabilities. Moreover, Ye et al. (2009) highlighted the advantages of the Gradient Boosted Tree model as one with high accuracy and ease of interpretation in classification. Therefore, it was determined that all of the aforementioned analytical techniques exhibited strengths in terms of their efficiency for data segmentation and complex forecasting. As a result, they were selected for this study.

However, according to a review of the literature on the application of these machine learning techniques to housing project prediction, Rico-Juan and de La Paz (2021) revealed that the Random Forest model demonstrated higher accuracy in predicting housing prices compared to other models. This is supported by the findings of Fan et al. (2006), who reported that the Decision Tree model also demonstrated strong accuracy in house price prediction. Furthermore, the study conducted by Lee et al. (2020) found that the Decision Tree model effectively correlated the factors affecting home

characterization. Moreover, Li et al. (2021) demonstrated the effectiveness of the Gradient Boosted Tree model in predicting factors that influence housing prices based on location. It was discovered that the factors affecting house price patterns include the distance to the city center, the presence of green scenery, and the population density of the area. This aligns with Truong et al. (2020), who found that the Gradient Boosted Tree model can effectively link location, zone, and demographic characteristics.

Furthermore, the ensemble classifier was also considered. A literature review by Ngoendee and Charoenruengkit (2021) showed an ensemble classifier of Random Forest and Decision Tree to investigate the relationship between locations and human behavior. The results showed that this combination of models could produce high accuracy in prediction. This is consistent with the findings of Ngoendee and Charoenruengkit (2021), who determined that the use of an ensemble model was the most effective and precise approach in data mining. The study by László and Ghous (2020) also concluded that the utilization of model integration techniques is an appropriate method for combining multiple models to obtain the most accurate and optimal predictions.

METHODOLOGY

The development of a machine learning model necessitates the consideration of key factors as inputs for the forecasting model. Due to the fact that each housing estate pattern in each country varies (Nilubon et al., 2019), the sub-factors of each group of factors must be taken into account when determining the type of housing estate development project. This information was extracted from the real estate market survey report in the Bangkok Metropolitan Region published by the Agency for Real Estate Affairs (2011, 2014, 2017, 2021). The selection of projects for the study was based on a comparison of average monthly sales rates data. The chosen projects were those with higher average monthly sales rates than the overall average for all sample projects. The types of residences studied included single-detached houses, semi-detached houses, and

townhouses, with a total of 59 projects - 22 single-detached, six semi-detached, and 31 townhouses.

After that, the researchers collected information including location details (distance from various amenities and facilities), nearby utilities, land appraisal price, land size, land use map, and project type, using market research reports and online sources. The collected data was verified, and any missing data was obtained from the project websites. The retrieved data were then analyzed through RapidMiner, a software recommended by László and Ghous (2020) for its high accuracy in data analysis. In total, seven models were analyzed, including: (1) Decision Tree (DT), (2) Random Forest (RF), (3) Gradient Boosted Tree (GBT), (4) an Ensemble Classifier of DT and RF, (5) an Ensemble Classifier of DT and GBT, (6) an Ensemble Classifier of RF and GBT, and (7) an Ensemble Classifier of DT, RF, and GBT. During prediction, the Ensemble Classifiers apply a majority vote criteria using the predictions of the sub-models as the input data (RapidMiner, 2023).

To filter the good-sales-performance data for the analyses, the 59 data results with higher-than-average sales rates were determined and divided into two groups using systematic sampling: 48 were used for model development, and 11 were used for model verification. The purpose of this division was to compare the obtained results and determine the technique that results in the most accurate model, based on the accuracy value. In addition, the researchers conducted an analysis of project patterns for all 59 projects to determine the factors affecting the specification of a housing development project of various types through descriptive statistics in accordance with the machine learning techniques.

RESULTS

In this section, the results of the data selection, which includes the details of the factors used as variables in the machine learning process, are presented. The obtained results from the study are as follows.

Proportion of Data Before and After Data Screening

The proportion of the data before and after the screening process are shown in Figure 3.

Figure 3 presents a classification of the data before and after the data screening process by eliminating the data with an average monthly sales rate lower than the average of all sample projects. The pre-screening data set consisted of a total of 179 data points. It was found that townhouses were the most numerous housing type, with a total of 100 projects (56%). Single houses made up 32% of the total with 58 projects, followed by semi-detached houses with 21 projects, or 12%. As to the post-screening data set of 59 data, townhouses were found to be the largest category of housing type, with 31 projects representing 53%, followed by 22 detached housing projects, 37%, and six semi-detached house projects, 10%.

Factors in Data Analysis

From the selection of 59 projects with average monthly sales rates higher than the average of all sample projects, which included single-detached houses, semi-detached houses, and townhouses, data were collected according to the principles identified in the literature review. The obtained details of the factors can be divided as in Table 1.

Table 1 presents the factors obtained from literature reviews, real estate market survey reports, and internet searches that are grouped into properties of the project and its location. These factors served as inputs in subsequent data analysis using machine learning techniques.

Machine Learning Models

Model Accuracy

After developing the model using data from housing projects with higher-than-average monthly sales rates through seven machine learning techniques, it was found that all acquired models produced accuracy, class recall, and class precision values, as shown in Table 2. The accuracy of a model can be calculated from the percentage of the total correct prediction cases to the total cases. In addition, the class precision of the model for a type of project can be calculated by the percentage of the correctly predicted cases to the total cases for the type of project, while the class recall of the model for a type of project can be calculated by the percentage of the detected cases to the total true cases for the type of project. These performance values can be calculated from a confusion matrix. An example of the calculations will be illustrated in detail in Section of Precision and Recall Values of Proposed Model.

Based on the accuracy values, it was determined that the Ensemble Classification of DT & GBT model achieved the highest accuracy of 90.91%. The other models, including Decision Tree, Random Forest, Ensemble Classification of DT & RF, Ensemble Classification of RF & GBT, and Ensemble Classification of DT & RF & GBT, produced accuracy values of 81.82%, and the Gradient Boosted Tree model had the lowest accuracy of 72.73%. Consequently, it can be inferred that all tested models have high accuracy. However, one limitation should be noted for the study, which was the limited number of datasets used in the model testing that led to high fluctuations in the recall and precision values. Particularly, the recall and precision of the model in the case of semi-detached houses, which had only one test dataset, produced only two values of 0.00% and 100.00%, as shown in Table 2.

Figure 3

Comparison of Proportions Before and After Data Screening.

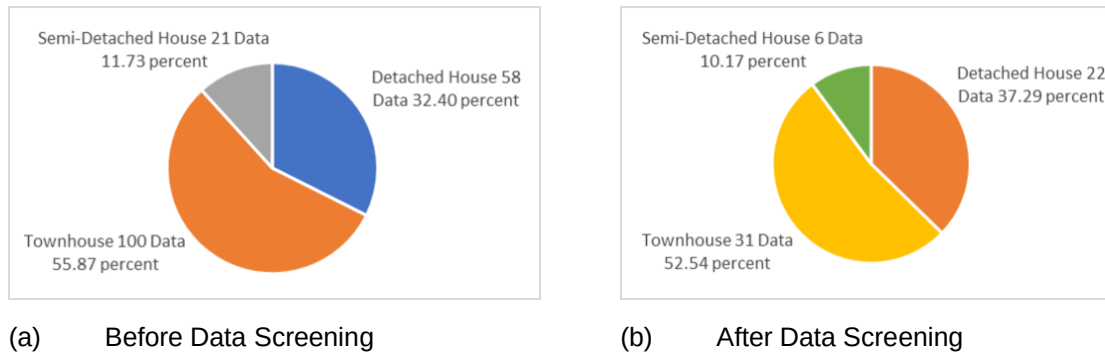


Table 1

Details of Factors Retrieved from Data Collection

Factor Types	Details
Project properties	- Entrepreneur type - Appraisal price of land in the project
Location properties	- Project zoning - Distance from expressway - Distance from sky train station - Distance from department store - Distance from public park - Distance from bus stop - Distance from market - Distance from hospital - Distance from main road - Distance from temple - Distance from petrol station

Table 2

Validity of Data Used in the Models

Model	Accuracy	Recall			Precision		
		SDH	TW	TH	SDH	TW	TH
Gradient Boosted Tree (GBT)	72.73%	75.00	0.00	83.33	75.00	0.00	71.43
Decision Tree (DT)	81.82%	75.00	100.00	83.33	100.00	100.00	83.33
Random Forest (RF)	81.82%	100.00	0.00	83.33	80.00	0.00	83.33
Ensembled Classification of DT & RF	81.82%	100.00	100.00	83.33	80.00	100.00	83.33
Ensembled Classification of RF & GBT	81.82%	100.00	0.00	83.33	80.00	0.00	83.33
Ensembled Classification of DT & RF & GBT	81.82%	100.00	100.00	83.33	80.00	100.00	83.33
Ensembled Classification of DT & GBT	90.91%	100.00	100.00	83.33	80.00	100.00	100.00

Note. SDH = Single-detached house, TW = Semi-detached house (twin house), TH = Townhouse

Proposed Model

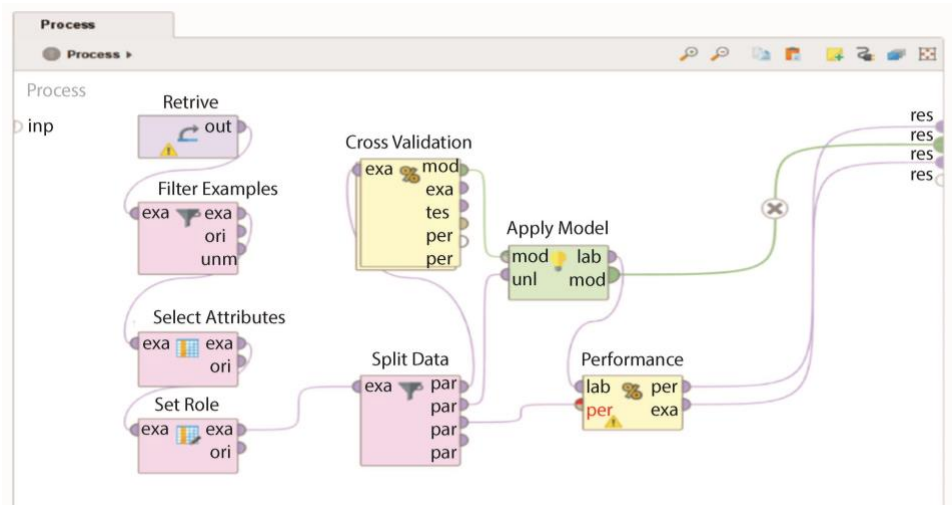
From Table 2, it can be concluded that the Ensemble Classification of DT & GBT technique (as illustrated in Figure 4) achieved the highest data accuracy of 90.91%. The model was acquired by combining the Decision Tree model (as shown in Figure 5), with the Gradient Boosted Tree, which consisted of 50 sub-model's sub-trees. It should be noted that there are five key criteria for the classification in the acquired Decision Tree model: Distance from main road, Distance from sky train station, Distance from bus stop, Distance from hospital, and Distance from shopping mall.

Precision and Recall Values of Proposed Model

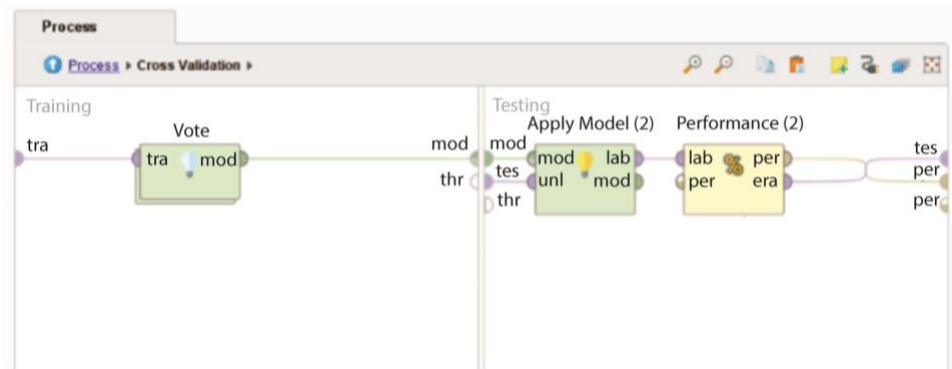
The proposed model, Ensembled Classification of Decision Tree (DT) & Gradient Boosted Tree (GBT) model, has Precision and Recall values shown in Table 2, where the details of how they can be calculated are shown in Table 3.

Table 3 provides details on the calculation of the Precision and Recall values of the proposed model. The Precision values for single-detached houses, semi-detached houses, and townhouses are 80.00%, 100.00%, and 100.00%, respectively, indicating the accuracy of the forecast for different housing project types. Meanwhile, the Recall values for single-detached houses, semi-detached houses, and townhouses are 100.00%, 100.00%, and 83.33% respectively, reflecting the ability to retrieve data for various housing project types.

Figure 4
Ensembled Classification Model

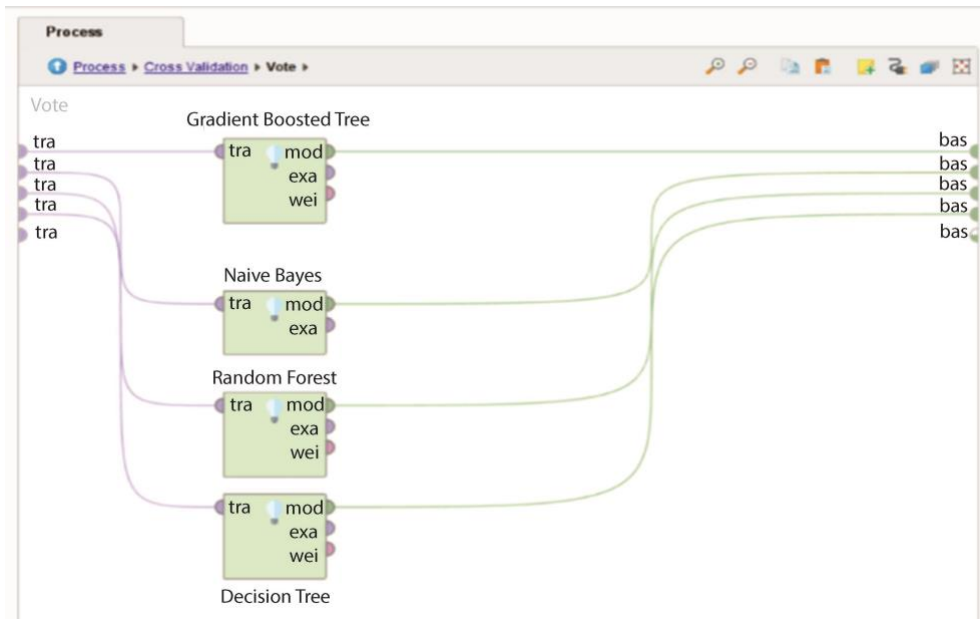


(a) Process Operators of Ensembled Classification Model



(b) Cross Validation Operator

Figure 4 (Continued)

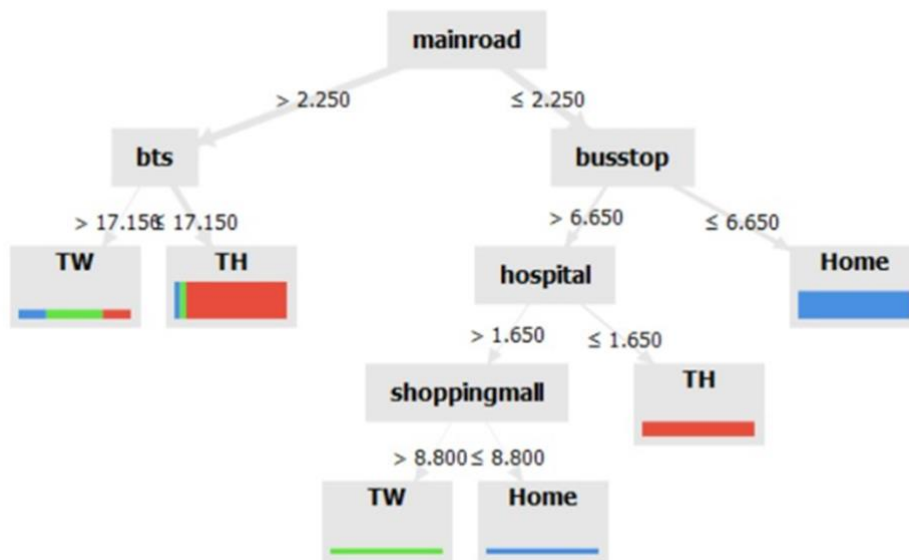


(c) Vote Operator

Note. From Vote by RapidMiner, 2023
(<https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/ensembles/vote.html>).
Copyright 2023 by RapidMiner.

Figure 5

Decision Tree Model



Note. From Vote by RapidMiner, 2023
(<https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/ensembles/vote.html>).
Copyright 2023 by RapidMiner.

RapidMiner. (2023). *Vote (apidMiner Studio Core)* [Computer software]. Rapidminer Documentation. Retrieved July 8, 2023, from <https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/ensembles/vote.html>

Table 3
Precision and Recall Values of Proposed Model

	True SDH	True TW	True TH	Class Precision
Predicted SDH	4	0	1	80.00%
Predicted TW	0	1	0	100.00%
Predicted TH	0	0	5	100.00%
Class Recall	100.00%	100.00%	83.33%	

Note. SDH = Single-detached house, TW = Semi-detached house (twin house), TH = Townhouse

DISCUSSION

The results of the analysis showed the effectiveness of using precise machine learning techniques for real estate problem analysis. This aligns with the findings of Afonso et al. (2019), who researched the forecasting of housing project sale prices to aid decision-making for entrepreneurs and residents. The study found that combining the Random Forest technique with Artificial Neural Networks, through combining models, resulted in better performance compared to using a single model technique, which is one of the most effective machine learning approaches.

The results retrieved from the Decision Tree model indicated that the factors considered for classifying the type of housing project included the proximity to a main road, train station, bus stop, hospital, and department store, etc., which is in line with the research by Sirithananonsakun and Tochaiwat (2018). It was found that (1) the land which is (1a) located within 2.25 km of a main road and within 6.65 km of a bus stop or (1b) located within 2.25 km of a main road and more than 6.65 km and 1.65 km from a bus stop and a hospital respectively, but within 8.80 km of a department store, is suitable for developing single-detached houses; (2) the land which is (2a) located more than 2.25 km from a main road and more than 17.15 km from a sky train station or (2b) located within 2.25 km of a main road but more than 6.65, 1.65 and 8.8 km from a bus stop, a hospital, and a department store, respectively, is suitable for semi-detached houses; and (3) the land which is (3a) located more than 2.25 km from a main road but within 17.15 km of a sky train station or (3b) located within 2.25 km and 1.65 km of a main road and a hospital,

respectively, but more than 6.65 km from a bus stop is suitable for developing townhouses.

The results indicated that the factors affecting project formulation include the location properties, represented by the distance from a bus stop and the distance from a park, as well as the project properties of land appraisal price. This aligns with the findings of Khan et al. (2017) and Rymarzak and Siemińska (2012), who noted that the distance to important places in each area plays a crucial role in determining the project style and, therefore, the location's value and selling price. Furthermore, the study supports the research by Li et al. (2021) that the value and potential of a location are dependent on the proximity to the city center and green spaces, such as parks.

Additionally, the analysis of data generated by machine learning techniques revealed that the Decision Tree and Random Forest models produced the highest accuracy for the single model cases (81.82% accuracy). However, the analysis further indicated that combining Decision Tree with Gradient Boosted Tree resulted in the most accurate model with higher accuracy (90.91%).

**CONCLUSION AND
RECOMMENDATIONS**

The results of the modeling performed using four machine learning techniques showed that a bus stop, local park, and land appraisal price were the factors affecting the determination of the project type, according to the decision tree. These findings can be used by entrepreneurs as a guide and recommendation as follows.

1. **Single-Detached House Project:** It is ideal for areas with high potential and value that have easy access to essential amenities and transportation hubs. If the location has moderate potential and value, it should still be situated if it is near to recreational areas, such as parks or green spaces.
2. **Townhouse Project:** This type of project is suitable for areas with moderate potential and value. Although it may be farther from certain key amenities, it should still have convenient access to transportation, such as bus stations. Additionally, proximity to green spaces, like parks, is important, even if the location potential is lower compared to more prime areas.
3. **Semi-detached House Project:** It is appropriate for areas with moderate potential and value. Although the proximity to important locations may be limited, the accessibility via convenient transportation is a critical factor to consider. Entrepreneurs should evaluate the distance from key transportation hubs, such as main roads or expressways, which may differ based on the specific context of each location.

However, it should be noted that this research has a limitation due to the limited sample size obtained from the data collection process, which had to acquire the project absorption rate data from market research reports. Entrepreneurs normally treat such data as confidential because it can be used in the marketing strategy formulation process of their competitors. Therefore, in applying the results from this research, entrepreneurs and stakeholders should supplement the findings with a thorough analysis of the context of each project. This limitation made the model performance indicators (accuracy, class precision and class recall) of all models seem to be in some specific numbers. For example, several acquired models have the same 81.82% accuracy. Researchers who are interested in doing similar research may try other approaches to cope with the overfitting problem that may occur. To illustrate, if the objective of the study is to identify the features that have impacts on the sales rate of the housing project, the researcher can use all data but measure the feature importance instead of filtering only high-sales-performance data, to keep more data for the model development process.

DISCLOSURE AND ACKNOWLEDGEMENT

Associate Professor Dr. Kongkoon Tochaiwat conceived the study and were responsible for the research design, data analysis and responsible for article preparation and submission.

Patcharida Seni Wong was responsible for data collection and data analysis

The research was supported by Thammasat University Research Fund, Contract No. TUFT 85/2564.

REFERENCES

Acheampong, R. A. (2018). Towards incorporating location choice into integrated land use and transport planning and policy: A multi-scale analysis of residential and job location choice behaviour. *Land Use Policy*, 78, 397–409. <https://doi.org/10.1016/j.landusepol.2018.07.007>

Afonso, B., Melo, L., Oliveira, W., Sousa, S., & Berton, L. (2019). Housing prices prediction with a deep learning and random forest ensemble. *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*, 16, 389–400. <https://doi.org/10.5753/eniac.2019.9300>

Agency for Real Estate Affairs. (2011). *An in-depth look at Bangkok and surrounding areas in the middle of 2011*. Agency for Real Estate Affairs.

Agency for Real Estate Affairs. (2014). *An in-depth look at Bangkok and surrounding areas in the middle of 2014*. Agency for Real Estate Affairs.

Agency for Real Estate Affairs. (2017). *An in-depth look at Bangkok and surrounding areas in the middle of 2017*. Agency for Real Estate Affairs.

Agency for Real Estate Affairs. (2021). *An in-depth look at Bangkok and surrounding areas in the middle of 2021*. Agency for Real Estate Affairs.

- Araloyin, F. M., & Olatoye, O. J. O. (2011). Factors affecting the provision of quality service in real estate agency in Lagos Metropolis, Nigeria. *International Journal of Business Administration*, 2(1), 71–79. <http://dx.doi.org/10.5430/ijba.v2n1p71>
- Baldominos, A., Blanco, I., Moreno, A.J., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying real estate opportunities using machine learning. *Applied Sciences*, 8(11), 1–23. <https://doi.org/10.3390/app8112321>
- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Big Data Thailand. (2020). *4 types of data segmentation (clustering)*. Retrieved April 25, 2022, from <https://bigdata.go.th/big-data--101/4-types-of-clustering/>
- Budhathoki, D. K. (2015). Quality management: An effective approach to success. *Journal of Nepalese Business Studies*, 9(1), 87–90.
- Ching, F. D. (2023). *Architecture: Form, space, and order* (4th ed.) John Wiley & Sons.
- DDproperty. (2020). *DDProperty Property Market Outlook 2020*. Retrieved January 25, 2024, from www.ddproperty.com/insights/ddproperty-property-market-outlook-2020-2-556.
- Egemen, M. (2022). Building construction clients' design consultant and contractor selection criteria versus post-occupancy satisfaction levels. *SAGE Open*, 12(2), 1–19. <https://doi.org/10.1177/21582440221089968>
- Elsinga, M., & Hoekstra, J. (2005). Homeownership and housing satisfaction. *Journal of Housing and the Built Environment*, 20, 401–424. <https://doi.org/10.1007/s10901-005-9023-4>
- Fan, G. Z., Ong, S. E., & Koh, H. C. (2006). Determinants of house price: A decision tree approach. *Urban Studies*, 43(12), 2301–2315. <https://doi.org/10.1080/00420980600990928>
- Gavankar, S. S., & Sawarkar, S. D. (2017). Eager decision tree. *IEEE 2017 2nd International Conference for Convergence in Technology (I2CT)* (pp. 837–840). <https://doi.org/10.1109/I2CT.2017.8226246>
- Habib, K. M. N., & Kockelman, K. (2008). Modeling choice of residential location and home type: Recent movers in Austin Texas. *87th Annual Meeting of the Transportation Research Board*. The University of Texas at Austin. www.caee.utexas.edu/prof/kockelman/public_html/TRB08HomeType%26Location.pdf
- Henilane, I. (2016). Housing concept and analysis of housing classification. *Baltic Journal of Real Estate Economics and Construction Management*, 4(1), 168–179. <http://dx.doi.org/10.1515/bjreecm-2016-0013>
- Kamglin, S., & Tochaiwat, K. (2014). Developer's decision-making process for real estate project type determination. *Proceedings of the 5th Built Environment Research Associates Conference*. Faculty of Architecture and Planning, Thammasat University.
- Khan, P. A. M., Azmi, A., Juhari, N., Khair, N., Daud, S. Z., & Rahman, T. (2017). Housing preference for first time home buyer in Malaysia. *International Journal of Real Estate Studies*, 11(2), 1–6.
- Koklic, M. K., & Vida, I. (2009). A strategic household purchase: Consumer house buying behavior. *Managing Global Transitions*, 7(1), 75–96.
- Klinchuanchun, P. (2022). *Industry outlook 2022-2024: Housing in BMR*. Krungsri Research. Retrieved August 12, 2023, from www.krungsri.com/en/research/industry/industry-outlook/real-estate/housing-in-bmr/io/housing-in-bmr-22

- Kung, E. (2022). Innovation and entrepreneurship in housing. *The Role of Innovation and Entrepreneurship in Economic Growth* (pp. 499–536). University of Chicago Press.
- László, K., & Ghous, H. (2020). Efficiency comparison of Python and RapidMiner. *Multidiszciplináris Tudományok*, 10(3), 212–220. <https://doi.org/10.35925/ji.multi.2020.3.26>
- Lee, J., Kim, G., & Choi, K. (2020). The relationship between childbirth, housing and socio-economic factors: The pattern analysis using decision tree. *Journal of Development and Communication Studies*, 21(2), 327–336.
- Li, L., & Wu, X. (2014). Housing price and entrepreneurship in China. *Journal of Comparative Economics*, 42(2), 436–449. <https://doi.org/10.1016/j.jce.2013.09.001>
- Li, S., Jiang, Y., Ke, S., Nie, K., & Wu, C. (2021). Understanding the effects of influential factors on housing prices by combining extreme gradient boosting and a hedonic price model (XGBoost-HPM). *Land*, 10(5), 1–15.. <https://doi.org/10.3390/land10050533>
- Luque, J. P., Ikromov, N., & Noseworthy, W. B. (2019). *Affordable housing development: Financial feasibility, tax increment financing and tax credits*. Springer International Publishing.
- Luttik, J. (2000). The value of trees, water and open spaces as reflected by house prices in the Netherlands. *Landscape and Urban Planning*, 48, 161–167. [http://dx.doi.org/10.1016/S0169-2046\(00\)00039-6](http://dx.doi.org/10.1016/S0169-2046(00)00039-6).
- Mahesh, B. (2020). Machine learning algorithms- A review. *International Journal of Science and Research*, 9(1), 381–386. <https://www.ijsr.net/archive/v9i1/ART20203995.pdf>
- Maoludyo, F. T., & Aprianingsih, A. (2015). Factors influencing consumer buying intention for housing unit in Depok. *Journal of Business and Management*, 4(4), 484–493.
- Mariadas, P. A., Abdullah, H., & Abdullah, N. (2019). Factors influencing the first home purchase decision of middle-income earners (M40) in Selangor, Malaysia. *Journal of Social Sciences and Humanities*, 16(1), 1–11. <http://journalarticle.ukm.my/15664/1/31705-97983-1-SM.pdf>
- Mete, M. O., & Yomralioglu, T. (2022). A hybrid approach for mass valuation of residential properties through Geographic Information Systems and Machine Learning Integration. *Geographical Analysis*, 5(4), 535–559. <https://doi.org/10.1111/gean.12350>
- Moghayedi, A., Awuzie, B., Omotayo, T., Le Jeune, K., Massyn, M., Ekpo, C. O., & Byron, P. (2021). A critical success factor framework for implementing sustainable innovative and affordable housing: A systematic review and bibliometric analysis. *Buildings*, 11(8), 1–31. <https://doi.org/10.3390/buildings11080317>
- Mulyano, Y., Rahadi, R. A., & Amaliah, U. (2020). Millennials housing preferences model in Jakarta. *European Journal of Business and Management Research*, 5(1), 1–9. <https://doi.org/10.24018/ejbmr.2020.5.1.240>
- Nakamura, H. (2019). Relationship among land price, entrepreneurship, the environment, economics, and social factors in the value assessment of Japanese cities. *Journal of Cleaner Production*, 217, 144–152. <https://doi.org/10.1016/j.jclepro.2019.01.201>
- Ngoendee, W., & Charoenruengkit, W. (2021). *Book recommendation with data mining using RapidMiner* [Master's thesis, Faculty of Science, Srinakharinwirot University]. <http://ir-thesis.swu.ac.th/dspace/handle/123456789/1232>

- Nilubon, P., Veerbeek, W., & Zevenbergen, C. (2019). Integrating climate adaptation into asset management planning: Assessing the adaptation potential and opportunities of an urban area in Bangkok. *International Journal of Water Resources Engineering*, 4(2), 50–65. <https://www.researchgate.net/publication/332292569>
- Nursoleh, N. (2022). Location analysis of interest in buying housing in South Tangerang city. *AKADEMIK: Jurnal Mahasiswa Ekonomi & Bisnis*, 2(1), 35–42. <https://doi.org/10.37481/jmeh.v2i1.247>
- Patel, B. R., & Rana, K. K. (2014). A survey on decision tree algorithm for classification. *International Journal of Engineering Development and Research*, 2(1), 1–5. <https://www.ijedr.org/papers/IJEDR1401001.pdf>
- Phe, H. H., & Wakely, P. (2000). Status, quality and the other trade-off: Towards a new theory of urban residential location. *Urban Studies*, 37(1), 7–35. <https://doi.org/10.1080/0042098002276>
- RapidMiner. (2023). *Vote (rapidminer studio core)* [Computer software]. Rapidminer Documentation. Retrieved July 8, 2023, from <https://docs.rapidminer.com/latest/studio/operator/s/modeling/predictive/ensembles/vote.html>
- Real Asset. (2021). *Should you buy or build a home?* Retrieved July 20, 2021, from www.realasset.co.th/living-consultant/detail/should-you-buy-or-build-a-home
- Real Estate Information Center. (2018). *The situation of the housing market in Bangkok-surrounding areas in the 3rd Quarter of 2018 and Trends Year 2019*. Retrieved November 2, 2021, from [www.reic.or.th/Upload/สถานการณ์อสังหาริมทรัพย์กรม-ปริมาตรไตรมาส3ปี2561-และแนวโน้มปี2562แก้ล่าสุด\(2\)_19356_1542375418_39809.pdf](http://www.reic.or.th/Upload/สถานการณ์อสังหาริมทรัพย์กรม-ปริมาตรไตรมาส3ปี2561-และแนวโน้มปี2562แก้ล่าสุด(2)_19356_1542375418_39809.pdf)
- Real Estate Information Center. (2022). *The situation of the housing market in Bangkok – post-COVID*. Retrieved February 20, 2022, from www.reic.or.th/News/RealEstate/455280
- Rico-Juan, J. R., & de La Paz, P. T. (2021). Machine learning with explainability or spatial hedonics tools? An analysis of the asking prices in the housing market in Alicante, Spain. *Expert Systems with Applications*, 171, 1–14. <https://doi.org/10.1016/j.eswa.2021.114590>
- Rissanen, T., Ermolaeva, L., Torkkeli, L., Ahi, A., & Saarenketo, S. (2020). The role of home market context in business model change in internationalizing SMEs. *European Business Review*, 32(2), 257–275. <https://doi.org/10.1108/EBR-09-2018-0168>
- Rymarzak, M., & Siemińska, E. (2012). Factors affecting the location of real estate. *Journal of Corporate Real Estate*, 14(4), 214–225. <https://doi.org/10.1108/JCRE-11-2012-0027>
- Shen, L., & Ross, S. (2021). Information value of property description: A machine learning approach. *Journal of Urban Economics*, 121, 1–49. <https://doi.org/10.1016/j.jue.2020.103299>
- Sirithananonsakun, N., & Tochaiwat, K. (2018). The location selection factors for housing development in Bangkok. *Built Environment Research Associates Conference [BERAC 9]* (pp. 532–540). Faculty of Architecture and Planning, Thammasat University. https://www.researchgate.net/publication/326261315_THE_LOCATION_SELECTION_FACTORS_FOR_HOUSING_DEVELOPMENT_IN_BANGKOK
- Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, 134, 93–101. <https://doi.org/10.1016/j.eswa.2019.05.028>
- Terra Media and Consulting. (2017). *3 important factors to consider before deciding on a location*. Retrieved January 25, 2024, from www.terrabbkk.com/articles/134597
- Tochaiwat, K. (2020). *Kanphatthana khrongkan chatsan* [Subdivision project development]. Thammasat Press.

- Tochaiwat, K., Kulintonprasert, W., Jiraprasertkun, C., & Yang, C.Y. (2018). Influence of master plan designs on housing project purchase decisions in Bangkok Metropolitan areas. *International Journal of Building, Urban, Interior and Landscape Technology*, 11, 43–54. <https://ph02.tci-thaijo.org/index.php/BUILT/article/view/167878>
- Zhang, F., Zhang, C., & Hudson, J. (2018). Housing conditions and life satisfaction in urban China. *Cities*, 81, 35–44. <https://doi.org/10.1016/j.cities.2018.03.012>
- Torbica, Ž. M., & Stroh, R. C. (2000). HOMBSAT- An instrument for measuring home-buyer satisfaction. *Quality Management Journal*, 7(4), 32–44. <https://doi.org/10.1080/10686967.2000.11918919>
- Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing price prediction via improved machine learning techniques. *Procedia Computer Science*, 174, 433–442. <https://doi.org/10.1016/j.procs.2020.06.111>
- Vandeviver, C., & Bernasco, W. (2020). Location, location, location: Effects of neighborhood and house attributes on burglars' target selection. *Journal of Quantitative Criminology*, 36, 779–821. <https://doi.org/10.1007/s10940-019-09431-y>
- Yang, F. J. (2018). An implementation of naive bayes classifier. *2018 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 301–306).. <https://doi.org/10.1109/CSCI46756.2018.00065>
- Ye, J., Chow, J. H., Chen, J., & Zheng, Z. (2009). Stochastic gradient boosted distributed decision trees. *Proceedings of the 18th ACM Conference on Information and Knowledge Management* (pp. 2061–2064). Association for Computing Machinery. <http://dx.doi.org/10.1145/1645953.1646301>
- Zeng, R. (2013). *Attributes influencing home buyers' purchase decisions: A quantitative study of the Wuhan Residential Housing Market* [Doctoral dissertation, Southern Cross University].