26TH ANNUAL PACIFIC RIM REAL ESTATE SOCIETY CONFERENCE
CANBERRA, AUSTRALIA 19TH – 22ND JANUARY 2020

HERITAGE PROPERTY VALUATION USING MACHINE LEARNING ALGORITHMS

JUNAINAH MOHAMAD
Faculty of Architecture, Planning and Surveying
Universiti Teknologi MARA Perak Branch, Seri Iskandar Campus, Malaysia

NUR SHAHIRAH JA'AFAR
Faculty of Architecture, Planning and Surveying
Universiti Teknologi MARA Shah Alam, Malaysia

SURIATINI ISMAIL
Faculty of Architecture and Ekistics
Universiti Malaysia Kelantan, Malaysia

ABSTRACT

The aim of this paper was to test the use of machine learning algorithms in predicting the price of heritage property. The original dataset consists of 311 prewar shophouse transacted from 2004 to 2018 in North-East of Penang Island, Malaysia. After the filtration process, only 248 units of prewar shophouse heritage property were available and valid to be used. We developed a heritage property price prediction model based on machine learning algorithms such as neural network, random forest, support vector machine, k*-nearest neighbors and linear regression and compare their predictive performance. The results indicate that the random forest algorithm performed better than the other models in predicting the price of heritage property.*

Keywords: heritage property, machine learning, neural network, random forest, support vector machine, *k*-nearest neighbors, linear regression

INTRODUCTION

Heritage property value can be used as part of an economic indicator. This underlines the importance of having an accurate value of heritage property for better-informed decision-making. Essentially, it is important to assess the heritage property value in order to a) acknowledge and respect the full worth of the heritage asset, b) appreciate the need for the maintenance and preservation of the heritage property, and c) assist in responding to calls for more accountability for the sustained use of the assets. The appropriate approach to valuation should produce an accurate value which is reliable and practical. The methods than can be used to estimate the value of heritage property are sale comparison method, cost method, contingent valuation method (CVM) and regression models. However, there has been no conclusive evidence as to what is the appropriate method of valuation for heritage property.

Previous study by the author has compared multiple regression analysis (MRA), rank transformation regression (RTR) and CVM in predicting the value of heritage property (Mohamad and Ismail, 2019). However, the idea of comparing between MRA, RTR and CVM needs to be viewed more thoroughly because these methods differ in functional forms and explanatory variables. Therefore, the aim of this paper is to investigate the potential of machine learning algorithms in predicting the value of heritage property as the method has same functional forms and explanatory variables with MRA and RTR.

LITERATURE REVIEW

Heritage property can be classified into two, for public use and for private use. In predicting the heritage property value as public goods, the term economic valuation has been used while for private heritage property the term market valuation or financial valuation has been used. The theory, method, and practice of economic valuation are well established especially in valuing Category I heritage property (for example, monuments museum). However, no studies have attempted to identify the most appropriate approach for valuing Category 2 heritage property (for example, prewar shop house). It is important to give attention to this type of heritage property as it contributes the most in terms of the number of transacted heritage property particularly in Malaysia. Table 1 shows the classification of shophouse heritage property, particularly in Georg Town, Pulau Pinang.

Table 1: Classification of shophouse heritage property

Grade	Description
Category 1	There are 82 buildings, gateways, cemeteries and sites in the George Town UNESCO World Heritage Site which are categorised as Category I. These buildings, monuments, objects and sites are important as they reflect the authenticity of the cultural landscape, largely contributing to the Outstanding Universal Values. They include: Buildings, monuments, objects and sites of exceptional interest Buildings and monuments declared as ancient and gazetted formerly under the Antiquities Act 1976 now under the National Heritage Act 2005 Buildings and monuments registered as National Heritage under the National Heritage Act 2005 The use of Category I buildings and sites should remain as originally intended, or be of similar use or nature of activity. Repairs carried out should use authentic, traditional ways of building methods and materials.
Category 2	Buildings, objects and sites of special interest that warrant every effort being made to preserve them. The majority of properties identified as Category II are shophouses, whilst other types of Category II items or objects include: Compounds, boundary walls, gateposts and gates, landscapes, enclaves, granite pathways and sites Historic street furniture such as granite posts and chains, fountains, lamp posts, post boxes, tramlines and trolley bus poles, fire hydrants and fire assurance plaques, granite pathway and engineering brick drains
Compatible Development	Infill - Existing empty land or temporary structure where compatible redevelopment is permitted.
	Replacement - Existing building without any significant value where sensitive redevelopment is permitted.

Source: Georg Town World Heritage Incorporated

Heritage Property Valuation

Valuation is a critical stage in the activities that relate to preservation and maintenance of cultural heritage, including built cultural heritage. However, there is little knowledge or lack of studies on how heritage properties are assessed (Esther, 2007). Also there are no standard definitions of the term heritage. The researchers also need to understand the difference between value and worth. In accounting, value is usually entered in a balance sheet. According to Sayce (2009), value is estimated using market evidence based on comparable transactions in relation to rents and capitalisation rates, to direct capital transactions or to the capitalisation of maintainable profits. Worth, on the other hand, may be calculated using a cash flow approach or it may take into account nonmonetary values. Such estimates may be critical to the owner in management decisions. Also, worth can be categorised as subjective because it is normally prepared for individual owners to enable them to manage their assets strategically. Under the current accounting principles, worth is not measured while value is. Normally. worth is used as a management tool, and the same goes to heritage property, whose value needs to be determined by value and worth. This is due to the following reasons: (1) the value of cultural heritage asset is subjective; (2) not many transactions take place (not in an active market); and (3) the value of cultural heritage is used in decision-making for preservation and maintenance. Each built cultural heritage has values that differ to various stakeholders, who can consist of the local authority, expert valuers, and economists. Basically, the estimated value can be useful for local authority for quit rent and maintenance purposes, whereas for valuers, the estimated value can be beneficial in terms of producing more reliable, valid, and practical values for dealing purposes, such as buying and selling. To the economists, estimated value can be beneficial for decision-making, whether to maintain, rebuild, or demolish the building. The difficulties in valuing heritage property are as follows;-

a) The complexity of the term is a fundamental part of why heritage property become difficult to be assessed. First, one needs to understand the classification of heritage property and type of values.
b) The current methods for assessing the impact and outcomes of heritage property value are increasingly being questioned, both in terms of methodologies and the results illuminate our understanding. If the methodology of measurement is not accurate, the results are inconclusive.
c) The limited availability of data might not facilitate us in understanding the valuable aspects of heritage value.

Machine Learning in Real Estate

For this section, articles on real estate forecasting using machine learning were identified through electronic resources such as Scopus and Web of Science. During the initial search, the keyword "("machine learning" AND "real estate" AND "price*")" was used and list of related articles where found which returned 45 articles. A selection criterion was finalized and every article was selected according to the selection criteria from 45 articles. The selection criterion are (i) real estate, (ii) price prediction/valuation, (iii) machine learning, (iv) articles in English, (v) indexed journal. 45 published studies were identified as part of the systematic search, after screening process, a final set of 16 (details for each study are presented in Table 2).

Table 2: Machine learning in real estate forecasting and valuation

No.	Author	Year	Country	Types of Property	Total Transaction	Supervised Machine Learning		The Best Prediction Model/Algorithm
						Regression	Classification	
1.	(Schernthanner *et al.*, 2011)	2011	Germany	Housing	74,098 units	Random Forest	nil	Random Forest
2.	(Oladunni and Sharma, 2015)	2015	America	Housing	135 unit	Linear Regression, Gradient Boosting	nil	Gradient Boosting
3.	(Park and Kwon, 2015)	2015	Virginia	Housing	5,359 records	Decision Trees, Ensemble	Naïve Bayesian	Ensemble

No.	Author	Year	Country	Types of Property	Total Transaction	Supervised Machine Learning		The Best Prediction Model/Algorithm
						Regression	Classification	
4.	(Crosby *et al.*, 2016)	2016	UK	Housing	12,000 transactions	Decision Tree, Random Forest	nil	Decision Tree
5.	(Oladunni, 2016)	2016	America	Housing	2,075 units	Principal Component Regression (PCR)	Support Vector Machine, K-nearest Neighbors	Principal Component Regression (PCR)
6.	(Nejad, Lu and Behbood, 2017)	2017	Australia	Apartment	1967 units	Ensemble, Decision Tree, Random Forest	nil	Random Forest
7.	(Trawi and Telec, 2017)	2017	Poland	Real estate	12,439 units	Linear Regression, Neural Networks, Decision Tree	nil	Decision Tree
8.	(Horino *et al.*, 2017)	2017	Japan	Apartment	6,320,631 posts	nil	Support Vector Machine	Support vector
9.	(Gu and Xu, 2017)	2017	China	Housing	253 units	Linear Price Model, 10.Gradient Bo11.osting	nil	Gradient Boosting
10.	(Di, Satari and Zakaria, 2017)	2017	India	Housing	21,000 units	Linear Regression, Multivariate regression, Polynomial Regression	nil	Mix all models
11.	(Kilibarda, 2018)	2018	Serbia	Apartment	7,407 units	Linear Regression, Random Forest, Principal Component Analysis (PCA)	nil	Random Forest
12.	(Ma *et al.*, 2018)	2018	Beijing	Warehouse	25,900 rental listings	Linear Regression, Random Forest, Gradient Boosting	nil	Random Forest
13.	(Varma *et al.*, 1936)	2018	Mumbai	Housing	nil	Linear Regression, Neural Network, Random Forest, Gradient Boosting	nil	Neural Network
14.	(Baldominos *et al.*, 2018)	2018	Spain	Housing	2,266 units	Support Vector Regression, Ensemble, Neural Network	K-Nearest Neighbors	Ensemble Regression Trees
15.	(Lee *et al.*, 2018)	2018	South Korea	Real Estate	77,532 units	Neural Network, Random Forest	nil	Random Forest
16.	(Medrano and Delgado, 2019)	2019	China	Housing	89 units	Linear Regression, Support Vector Regression, Neural Network	K-Nearest Neighbors,	Support Vector Regression

Machine learning algorithms allow the data to make decisions based on its previous experience (Dutton and Conroy, 1997). In the past decade, machine learning has been used in many research areas and its diversity has attracted the use of the algorithms for different applications. In the past few years, researchers have started to use machine learning algorithms in real estate forecasting analysis. However, according to Table 2, the use of machine learning techniques in Malaysia real estate market is still undiscovered yet. Therefore, this paper aims to test the potential of machine learning in real estate valuation by taking heritage property as a case study.

Previous study by Phan (2018) stated that the use of machine learning in real estate market can be divided into two, which are trends in forecasting the house price index and house price valuation. In predicting the house price index, the author used vector auto regression model while the author used support vector machine for house price valuation. The machine learning can be grouped into two which are supervised and unsupervised (Ng and Deisenroth, 2015; Kaytan and Aydilek, 2017). Figure 1 shows the types of machine learning task and their algorithms. The most common machine learning task used for real estate is supervised learning, and the most common machine learning algorithms for real estate are shown in Table 2.

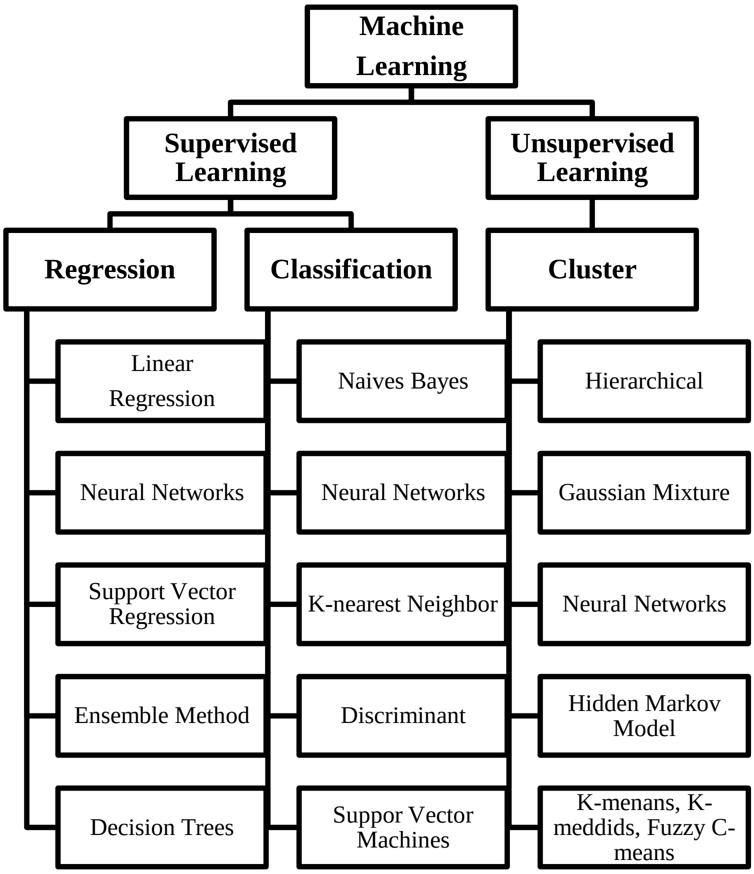


Figure 1:Types of machine learning task and their algorithms

Machine Learning Algorithms Used in This Study

The following subsection describe the neural network (NN), random forest (RF), support vector machine (SVM), k-nearest neighbors (kNN) and linear regression (LR) algorithms.

Neural Network

A neural network (NN) is a set of N neurons used to estimate non-linear functions of two or more neurons. Each neuron represents a parametrized non-linear function. An input variable value is entered into a neuron and depending on the function *f* the latter represents (normally with an associated parameter as a weight factor), an output value is obtained that may in turn be the input value for some other neuron. The function *f* is known as an activation function given that it may be activated depending on the input value. In this study we use a single hidden layer neural network with three layers (Ripley and Hjort, 1996). The first one is the input layer in which the neurons receive the dependent variable values and the third one is the output layer in which a neuron receives the linear combinations of the values of the second, or hidden, layer. This latter receives the input neuron values and from this information calculates the weights of the activation functions so as to optimize an output criterion. For this study the criterion is the minimum mean square error (MMSE). Various sizes of hidden-layer networks were tested with a view to maintaining the simplicity of the network without sacrificing performance. The size finally decided upon was 8 neurons. The weight decay is 0.1, which avoids problems of overfitting and improves the generalization of the network (Hertz *et al.*, 1991).

Random Forest

The RF algorithm is a method of bagging trees (Breiman, 1996). Under it, multiple unpruned classification trees are trained through iteration of samples without replacement of the original data set. Each tree classifies the instances individually and the forest as a whole then chooses the classifications having the most individual votes (over all of the trees) (Breiman, 2001). A key characteristic of this method is that the trained trees do not depend on the trees trained previously given that at each iteration a bootstrap sample of the data set is used. In this study, RF is used with regression trees given that the dependent variable is continuous. The procedure is the same as that for training classification trees, the sole difference being that in regression trees the leaves predict the actual number instead of a class. In this context the partition criterion is not entropy but rather the minimum root-mean-square-error (RMSE). The parameter found to generate the least RMSE was 10 variables at each growth of the tree.

Support Vector Machine

A support vector machine is a supervised classification algorithm that searches for a hyperplane separating data into classes. The separation is identified by maximizing the distance between the hyperplane and the region that defines the limit of each class. The SVM transforms the training data using a function known as kernel that maps the data to a higher-dimensional space, thus permitting a better hyperplane separator to be found. In the process of maximizing the distance, SVM assigns a cost or penalty to the classification of an instance in the wrong class and then minimizes these errors. Strictly speaking, the version of SVM utilized in this study is known as Support Vector Regression, which algorithm is the same as SVM except that the cost function is modified for use with a distance measure, in this case the conventional RMSE (Smola and Schölkopf, 2004). To build the SVM model we performed a set of simulations with a variety of parameter combinations, in particular the gamma and cost parameters. The least RMSE was obtained with equal to 0.1 and a cost of 10.

***k*-Nearest Network**

K-Nearest Neighbors (KNN) is a standard machine-learning method that has been extended to large-scale data mining efforts. The idea is that one uses a large amount of training data, where each data point is characterized by a set of variables. Conceptually, each point is plotted in a high-dimensional space, where each axis in the space corresponds to an individual variable. When we have a new (test) data point, we want to find out the K nearest neighbors that are closest (ie, most "similar" to it). The number K is typically chosen as the square root of N, the total number of points in the training data set. (Thus, if N is 400, K = 20).

KNN is conceptually simple and has the advantage of being nonparametric. That is, the method can be used even when the variables are categorical—though if you are using numeric variables in the mix, it is best to standardize them to eliminate differences in scale. The challenge is that when the number of data points is very large (eg, an online bookseller has millions of books), special methods must be employed to rapidly search the space and find the “most similar” items.
Usually, some form of precomputation is employed for example, indexing. In addition, rather than using all the data points, selected data points that are representative of individual clusters (“prototypes”) may be used to facilitate the search against a new item, and then the precomputed neighbors of the most similar prototype are also displayed. Similarly, attempting to reduce the number of dimensions with a method like SVD/LSI and then plotting the data points in the reduced variable space may result in significant gains in performance.

Linear Regression

Linear regression Varma *et al.*, (1936), is one of the well-known algorithms, also known as ordinary least squares (OLS). Consist of two types such simple linear regression (SLR) and multiple linear regressions (MLR). Simple Linear Regression is characterized by one independent variable while Multiple Linear Regression is characterized by more than one independent variable. Linear regression is used to estimate real world values like cost of houses, number of calls, total sales etc. Linear equation Y= a *X + b.

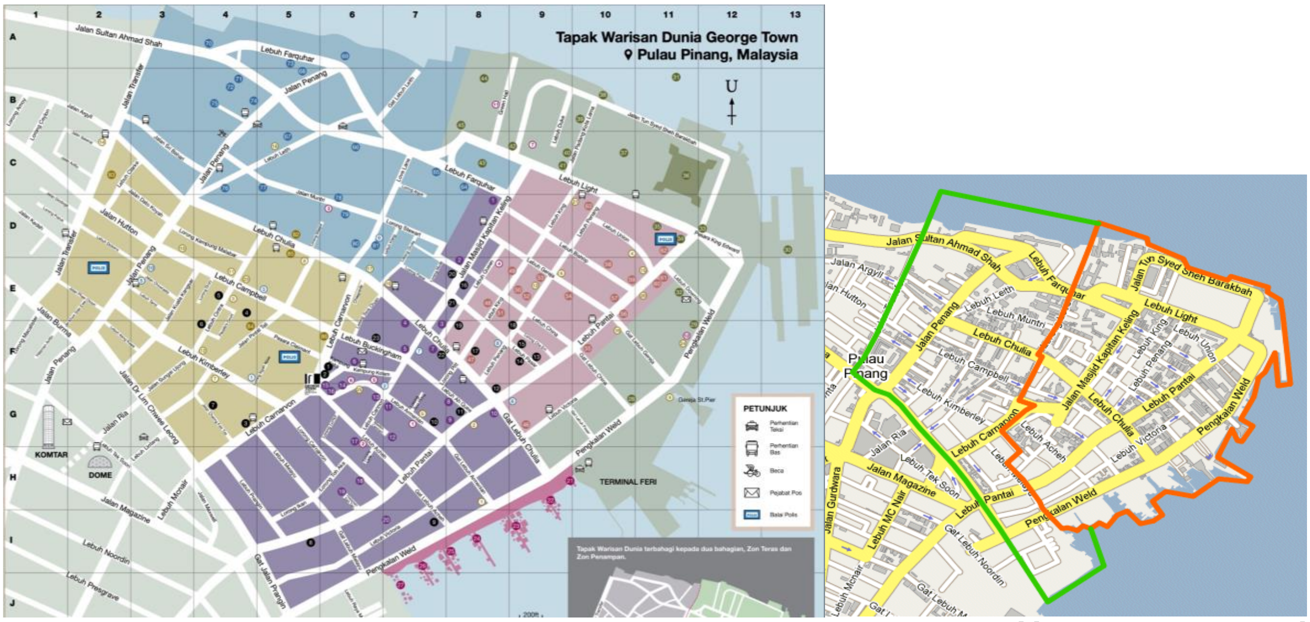
Calibration and Validation of Models and Performance Measures

To calibrate and validate the models generated by NN, RF, SVM, kNN and LR algorithms, training and validation data sets were tuned with a 10-fold cross validation. The performance measures used to evaluate the models were root-mean-squared-error (RMSE), mean-absolute-error (MAE) and the Pearson coefficient of determination (R^2). Note, the following observations regarding the three measures:

1) RMSE is an accuracy measure often used to compare the sample standard deviation of the differences between observed and predicted values. It also indicates the aggregate size of the errors in a model's predictions and is thus a measure of predictive power,
2) MAE gives the average of the squared errors. It is similar to RMSE except that the differences between the observed and predicted values are not squared and,
3) R^2 indicates how well the models fit the training and validation data sets. It measures the proportion of the variance explained by the models. The closer this proportion is to 1, the better.

RESEARCH METHODOLOGY

George Town, the capital city of the Malaysian state of Penang, is located at North-Eastern tip of Penang Island. It is Malaysian’s second largest city. The historical core of George Town has been inscribed as a UNESCO World Heritage Site since 2008. This paper focusing on heritage property valuation consist of prewar shophouse located at inner city of George Town, as shown in Figure **2**. The areas are divided by two zone which are core zone (orange line) and buffer zone (green line).



Source: Georg Town World Heritage Incorporated

Figure 2: The inner city of George Town

The secondary data on property transactions for this paper were collected in digital form from National Property Information Centre (NAPIC), Malaysia. The data contained record of prewar shophouse transaction in Penang Island, Malaysia from 2004 to 2018. The registered sale price was the actual price paid for the prewar shop house. Thus, the price data used in this study was transaction price. However, during filtration process, only arm's length transaction is considered. Other variables used for training the machine learning are road, zone, number of storey, year of transaction and lot size. Table 3 shows the filtering process of the original set of data from 2004 to 2018 in which 248 observations (prewar shop house) remained for this study. The data were examined for completeness and usefulness to develop the training machine learning algorithms. More information about the data and the range of the values is available in Table 4.

Table 3: A record of data cleaning process

No	Notes	Number of records left
1.	Original data from 2004 to 2018 for prewar shophouse in Penang Island, Malaysia from NAPIC	3121
2.	Excluding property not in core zone and buffer zone (located at North-East, Penang Island)	311
3.	Excluding share	260
4.	Excluding lot size	253
5.	Excluding number of storey – Final Data	248

Table 4: Descriptive statistics of final dataset

No.	Label/Code	Definition	Type of variables	Min	Max	Mean	Std. Dev
1.	Price	Transaction price (Ringgit Malaysia)	Nominal	38000	7500000	1257059.75	1198075.98
2.	Road	Road name/location of the property according to road name	Nominal	n.a	n.a	n.a	n.a

No.	Label/Code	Definition	Type of variables	Min	Max	Mean	Std. Dev
3.	Zone	Location of the property according to UNESCO heritage zone which are core zone and buffer zone (1=core zone, 0=buffer zone	Nominal	n.a	n.a	n.a	n.a
4.	Storey	Number of storey	Numeric	1	3	2.016	0.201
5.	Year	Year of transaction from 2004 to 2018	Numeric	2004	2018	2010	3.1
6.	Lot Size	Lot size	Numeric	33	1408	219.685	198.501

RESULTS AND DISCUSSION

Before developing the machine learning models, it is important to check the mutual effect of different variables to be used in constructing real estate models. It is well known among the real estate modelling researchers that multicollinearity between two independent variables is not a good thing. If the results show no relationships between explanatory variables (no correlation), they would be said to be statistically independent to another. If the variables are highly correlated, it will lead to unreliable and unstable estimates of regression coefficients (Brooks and Tsolacos, 2010). Table 5 shows the correlation indices of the original variables in order to indicate the multicollinearity among them. The results reveal that there are no variables that have collinearity index above 0.8.

Table 5: Correlation indices

		Storey	Year	Lot Size	Price
Storey	Pearson Correlation	1	-.024	-.030	.198**
	Sig. (2-tailed)		.709	.637	.002
	N	248	248	248	248
Year	Pearson Correlation	-.024	1	-.137*	.384**
	Sig. (2-tailed)	.709		.031	.000
	N	248	248	248	248
Lot Size	Pearson Correlation	-.030	-.137*	1	.574**
	Sig. (2-tailed)	.637	.031		.000
	N	248	248	248	248
Price	Pearson Correlation	.198**	.384**	.574**	1
	Sig. (2-tailed)	.002	.000	.000	
	N	248	248	248	248
**. Correlation is significant at the 0.01 level (2-tailed).					
*. Correlation is significant at the 0.05 level (2-tailed).					

This section compares the predictive performance of heritage price prediction model based on machine learning algorithms which are NN, RF, SVM, kNN and LR. The performance of the five algorithms in predicting the heritage property values in North-East Penang Island is indicated by the results shown in Table 6. For this task, we used the WEKA software, a knowledge analysis suite developed by the University of Waikato. Each of the model was tuned with a 10-fold cross validation and the error rates shown in Table 6. Based on the result of prewar shop houses, RF is the best algorithm in predicting the price of prewar shophouse with the least MAE and RMSE, and highest R^2 values.

Table 6: Performance indicators for each algorithm

Algorithm	NN	**RF**	SVMreg	kNN	LR
RMSE	950789.487	**628818.6907**	903517.1624	831592.7385	780603.1305
MAE	671700.9097	**368139.7289**	516764.7145	496345.1694	529845.0429
R^2	72.64%	**85.69%**	66.65%	73.2%	75.84%

CONCLUSIONS

The previous study on heritage property valuation by the author has focused on stated preference (CVM) and revealed preference (regression model) in the attempt to find appropriate model. There has been no evidence of the use of machine learning in heritage property studies (as shown in Table 2) despite its widespread application is other fields. This paper has presented a continued effort of finding appropriate valuation model for heritage property by focusing on machine learning algorithms.

This study has presented the empirical results based on 248 prewar shop houses in North-East, Penang Island, Malaysia using machine learning techniques. Several machine learning algorithms have been used to develop prediction models for prewar shophouse in North-East, Penang Island. Five different supervised machine learning algorithms used are NN, RF, SVM, kNN and LR. The findings show that, the RF model has produced the best predictions of prewar shop houses in North-East, Penang Island, Malaysia.

Nonetheless, this early study has some limitations which future research could examine further. Firstly, the variables used are only limited to those that have been made available in the databased provided by NAPIC. In future, the author aims to enrich the data by adding the location and historical characteristics which are important to heritage property. This could possibly improve the model with better predictions. Secondly, no hold-out samples have been used. This is important because in-sample performance may overstate the performance, especially certain machine learning algorithms like RF as mentioned by Mullainathan and Spiess (2017). Thirdly, this paper has not discussed the non-linearity characteristic of real estate market which Shimizu, Karato and Nishimura, (2014) reported that the non-linear models will increase the accuracy of the models.

In conclusion, this paper has revealed the potential of machine learning in predicting the price of heritage property by highlighting the superiority of RF algorithm. Our next endeavour will be to compare between machine learning algorithms, MRA, RTR and CVM.

ACKNOWLEDGEMENT

The authors express gratitude to the National Property Information Centre (NAPIC), Malaysia for providing the data and the Ministry of Education Malaysia, Higher Education for funding the research (FRGS: FRGS/1/2018/WAB03/UiTM/03/1). Our gratitude also goes to the anonymous and suggestions on the early draft of this paper.

REFERENCES

Baldominos, A. *et al.* (2018) ‘applied sciences Identifying Real Estate Opportunities Using Machine Learning’. doi: 10.3390/app8112321.

Breiman, L. (1996) ‘Bagging predictors’, *Machine learning*. Springer, 24(2), pp. 123–140.

Breiman, L. (2001) ‘Random forests’, *Machine learning*. Springer, 45(1), pp. 5–32.

Brooks, C. and Tsolacos, S. (2010) *Real estate modelling and forecasting*. Cambridge University Press.

Crosby, H. *et al.* (2016) ‘A Spatio-Temporal , Gaussian Process Regression , Real-Estate Price Predictor’, pp. 3–6.

Di, N. F. M., Satari, S. Z. and Zakaria, R. (2017) ‘Real estate value prediction using multivariate regression models Real estate value prediction using multivariate regression models’. doi: 10.1088/1757-899X/263/4/042098.

Dutton, D. M. and Conroy, G. V (1997) ‘A review of machine learning’, *The knowledge engineering review*. Cambridge University Press, 12(4), pp. 341–367.

Gu, G. and Xu, B. (2017) ‘Housing Market Hedonic Price Study Based on Boosting Regression Tree’, 21(6).

Hertz, J. *et al.* (1991) ‘Introduction to the theory of neural computation’, *Physics Today*, 44, p. 70.

Horino, H. *et al.* (2017) ‘Development of an Entropy-Based Feature Selection Method and Analysis of Online Reviews on Real Estate’, pp. 2351–2355.

Kaytan, M. and Aydilek, İ. B. (2017) ‘A review on machine learning tools’, in *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*. IEEE, pp. 1–4.

Kilibarda, M. (2018) ‘Estimating the Performance of Random Forest versus Multiple Regression for Predicting Prices of the Apartments’, (Ml). doi: 10.3390/ijgi7050168.

Lee, W. *et al.* (2018) ‘Machine Learning based Prediction of The Value of Buildings’, *KSII Transactions on Internet and Information Systems*, 12(8), pp. 3966–3991.

Ma, Y. *et al.* (2018) ‘Estimating Warehouse Rental Price using Machine Learning Techniques’, 13(April), pp. 235–250.

Medrano, C. and Delgado, J. (2019) ‘Estimation of missing prices in real-estate market agent-based simulations with machine learning and dimensionality reduction methods’, 7. doi: 10.1007/s00521-018-3938-7.

Mohamad, J. and Ismail, S. (2019) ‘Capabilities Of Revealed Preference Method For Heritage Property Valuation’, *Planning Malaysia Journal*, 17(9).

Mullainathan, S. and Spiess, J. (2017) ‘Machine learning: an applied econometric approach’, *Journal of Economic Perspectives*, 31(2), pp. 87–106.

Nejad, M. Z., Lu, J. and Behbood, V. (2017) ‘Applying Dynamic Bayesian Tree in Property Sales Price Estimation’.

Ng, A. and Deisenroth, M. (2015) ‘Machine learning for a London housing price prediction mobile application’, in *Imperial College London*.

Oladunni, T. (2016) ‘Hedonic Housing Theory – A Machine Learning Investigation Hedonic Housing Theory – A Machine Learning Investigation’, (December). doi: 10.1109/ICMLA.2016.0092.

Oladunni, T. and Sharma, S. (2015) ‘Predictive Real Estate Multiple Listing System Using MVC Architecture and Linear Regression 1 Introduction’.

Park, B. and Kwon, J. (2015) ‘Expert Systems with Applications Using machine learning algorithms for housing price prediction : The case of Fairfax County , Virginia housing data’, *EXPERT SYSTEMS WITH APPLICATIONS*. Elsevier Ltd, 42(6), pp. 2928–2934. doi: 10.1016/j.eswa.2014.11.040.

Phan, T. D. (2018) ‘Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia’, in *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*. IEEE, pp. 35–42.

Ripley, B. D. and Hjort, N. L. (1996) *Pattern recognition and neural networks*. Cambridge university press.

Sayce, S. (2009) *Valuing Heritage Assets, Examining the Case for the Valuationof Heritage Assets*. Edited by RICS and H. M. Treasury. Kingston University London.

Schernthanner, H. *et al.* (2011) 'Spatial modeling and geovisualization of rental prices for real estate portals'.

Shimizu, C., Karato, K. and Nishimura, K. (2014) 'Nonlinearity of housing price structure: Assessment of three approaches to nonlinearity in the previously owned condominium market of Tokyo', *International Journal of Housing Markets and Analysis*. Emerald Group Publishing Limited, 7(4), pp. 459–488.

Smola, A. J. and Schölkopf, B. (2004) 'A tutorial on support vector regression', *Statistics and computing*. Springer, 14(3), pp. 199–222.

Trawi, B. and Telec, Z. (2017) 'Comparison of Expert Algorithms with Machine Learning Models for Real Estate Appraisal'.

Varma, A. *et al.* (1936) 'House Price Prediction Using Machine Learning And Neural Networks', *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE, pp. 1936–1939.

Yung, H. K., & 容曉君. (2007) *Architectural heritage conservation in Hong Kong: an empirical analysis, HKU Theses Online (HKUTO)*. The University of Hong Kong (Pokfulam, Hong Kong). doi: 10.5353/th_b3893485.

Email contact: mjunainah@uitm.edu.my / mjunainah@gmail.com