UNIVERSITI TEKNOLOGI MARA
Cawangan Perak
Kampus Seri Iskandar

e-Proceeding

v-GOGREEN2020

29 - 30 SEPT

VIRTUAL GO-GREEN: CONFERENCE & PUBLICATION

"SUSTAINABLE ENVIRONMENT, RESILIENCE AND SOCIAL WELL-BEING"

Organiser :
Research, Industrial Linkages, Community & Alumni Network (PJIM&A)

Co-organiser :
Faculty of Architecture, Planning and Surveying (FSPU) & Centre for Post Graduate Studies (CGS)

Publication Date : 22nd February 2021

Virtual Go-Green Conference and Publication 2020

UNIVERSITI TEKNOLOGI MARA, PERAK BRANCH
February 2021

Main Committee

Virtual Go-Green Conference and Publication 2020

Advisor 1	*: Prof Sr Dr Md Yusof Hamid, AMP*
Advisor 2	*: Assoc Prof Dr Nur Hisham Ibrahim*
Chairman	*: Sr Dr Asmalia Che Ahmad*
Co-Chairman	*: 1. Sr Dr Yuhainis Abdul Talib*
	2. Sr Dr Haryati Mohd Isa
Treasurer	*: Mohamad Haizam Mohamed Saraf*
Secretary	*: Noorliza Musa*
Head of v-Conference	*: Sr Dr Nor Suzila Lop*
Head of e-Proceeding	*: Dr Junainah Mohamad*
Head of Scopus Indexed Journal Planning Malaysia Journal (PMJ)	*: Assoc Prof Gs Dr Mohd Fadzil Abdul Rashid*
Head of Scopus Indexed Journal Malaysian Construction Research Journal (MCRJ)	*: Sr Dr Natasha Khalil*
Head of Paper Reviewer	: *Dr Asniza Hamimi Abdul Tharim*

Committee Members

Virtual Go-Green Conference and Publication 2020

E-Proceeding Paper Reviewer

Noraini Md Zain
Shafikah Saharuddin
Nur Fatiha Mohamed Yusof
Farrah Rina Mohd Roshdi

E-Proceeding Formatting

Nurulanis ahmad @ Mohamed
Jannatun Naemah Binti Ismam
Najma Binti Azman

E-Proceeding Language Reviewer

Dr Hjh Shazila Abdullah
Dr Daljeet Singh Sedhu A/L Janah Singh
Zarlina Mohd Zamari
Dr Mary Thomas
Iza Faradiba Mohd Patel
Farahidatul Akmar Awaludin
Wan Faridatul Akma Wan Mohd Rashdi
Jeyamahla Veeravagu
Wan Nurul Fatihah Wan Ismail
Nazirul Mubin Mohd Noor
Noor Aileen Ibrahim
Noraini Johari
Dr Hajah Norakmarwati Ishak

Virtual Conference

Norazlin Mat Salleh	*Registration*
Shahela Mamter	*Auditor*
Mohd Esham Mamat	*Auditor*
Noor Anisah Abdullah @ Dolah	*Auditor*
Mohamad Tajudin Saidin	*Certificate & Conference Kit*
Fairiz Miza Yob Zain	*Logistic*
Mohd Firdaus Zainuddin	*Logistic*
Farah Hanna Ahmad Fuad	*Promotion & Publicity*
Mohamad Shahin Shahdan	*Promotion & Publicity*
Mohd Asrul Hassin	*Liason Officer*

Organiser:
Research, Industrial Linkage Community and Alumni Network Office (PJIM&A)
Universiti Teknologi MARA, Perak Branch, Seri Iskandar.
Malaysia

Co-Organiser:
Faculty of Architecture, Planning and Surveying (FSPU)
and,
Centre for Post Graduate Studies (CGS)
Universiti Teknologi MARA, Perak Branch, Seri Iskandar.
Malaysia

e ISBN 978-967-2920-06-9

TESTING THE USE OF MACHINE LEARNING FOR HERITAGE PROPERTY VALUATION

Junainah Mohamad[1], Nur Shahirah Ja'afar[2] and Suriatini Ismail[3]
[1]Department of Estate Management, Faculty of Architecture, Planning and Surveying, Universiti Teknologi MARA, Perak Branch, Seri Iskandar Campus, Seri Iskandar, 32610 Perak, Malaysia
[2]Department of Estate Management, Faculty of Architecture, Planning and Surveying, Universiti Teknologi MARA Shah Alam, Shah Alam, 40450 Selangor, Malaysia
[3]Faculty of Architecture and Ekistics, University Malaysia Kelantan, Malaysia

Abstract

Recently, the use of machine learning is gaining ground and it holds great promise for real estate valuation. However, the application of machine learning in heritage property valuation has limited adoption. Therefore, this paper aims to demonstrate the potential use of machine learning in heritage property valuation. The original dataset consists of 311 prewar shophouses transacted from 2004 to 2018 at North-East of Penang Island, Malaysia. After the filtering process, only 137 units of pre war shophouse heritage property were available and valid to be used. Several machine learning algorithms have been developed and tested, including random forest regressor, decision tree regressor, lasso, ridge and, linear regression. The results indicate that random forest regressor is the best machine learning algorithms and can be used for heritage property valuation.

Keywords: *heritage property valuation; machine learning; prewar shophouse*

1.0 INTRODUCTION

Property valuations can be challenging as there are so many factors that influence price. Current appraisal techniques are often based on previous sale price, it fails to take into account factors like building size, land size, amenities, local neighborhood and others that have an impact on a property value. This paper focuses on heritage property as the subject property to be valued. The values of heritage property can be used as part of an economic indicator. This indicates the importance of having an accurate value of heritage property for better-informed decision-making. It is important to assess the value of heritage property 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 that 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. Due to pattern recognition abilities, machine learning is seen to have troughs in property valuation. Machine learning software can greatly increase accuracy of price prediction. Instead of current valuation techniques like using previous sale data, machine learning could process exponentially more data points that have an impact on property values. Thus, the aim of this study is to test the applicability of machine learning in assessing the value of heritage property.

2.0 LITERATURE REVIEW

Valuation is a critical stage in the activities that relate to preservation and maintenance of cultural heritage, including built cultural heritage. However, there is a lack of studies on how heritage properties are being assessed (Yung, 2013). Moreover, there are also 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 non-monetary values. Such estimates may be critical to the owner in management decisions. Worth can also 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 assets 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.

Machine learning has become popular programming among researchers in predicting the price of property. Previous study by Phan (2018) stated that the use of machine learning in the real estate market can be divided into two, which are trends in forecasting the house price index and house price valuation. The machine learning can be grouped into two which are supervised and unsupervised (Kaytan & Aydilek, 2017; Ng & Deisenroth, 2015). Within these categories, the most commonly been used is supervised learning algorithms (Horino et al., 2017). The following describes the machine learning algorithms used in this study which are random forest regressor, decision tree regressor, lasso, ridge and linear regression. Random forest is a supervised classification algorithm, but it is still can be used for both classification and regression kind of problems, it is a trademark term for an ensemble of decision trees. In Random forest, we have a collection of decision trees, known as “Forest”. To classify a new object based on attributes, each tree gives a classification and we say the tree “votes” for that class. The forest chooses the classification having the most and highest votes. Random forest can handle missing values but it won’t over fit the model even data have more number of trees in the forest (Kilibarda, 2018). Decision tree, it is a supervised learning algorithm that covers both classification and regression and mostly used for classification problems. In decision analysis, a decision tree is used to visually and explicitly represent decisions and decision making. It uses a tree-like model of decisions. A decision tree is drawn with its root at the top and branches at the bottom. In the image, the bold text represents a condition/internal node, based on which the tree splits into branches/ edges. The branch end that does not split anymore is the decision/leaf. Lasso is known as L1 regularization technique because it is one of the powerful in regression, which works by reducing the error between predicted and actual observations. Last but not least is ridge regression, it is also a linear model for regression and it applies the same formula used in predictions using OLS (Muller & Guido, 2017). This algorithm is able to fit an additional constraint namely called regularization during training data. Ridge regression is known as L2 regularization, this regularization is to avoid overfitting during training. Linear regression (Varma, Sarma, Doshi, & Nair, 2018), 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 $Y= a *X + b$.

3.0 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 focuses on heritage property valuation consisting of pre-war shophouses located in the inner city of GeorgeTown, as shown in Figure 1. The areas are divided by two zone which are core zone (red line) and buffer zone (blue line).

Figure 1: Image of core zone and buffer zone in george town penang island

The secondary data on property transactions for this paper were collected in digital form from National Property Information Centre (NAPIC), Malaysia. The data contained records of pre war shophouse transactions in Penang Island, Malaysia from 2004 to 2018. The registered sale price was the actual price paid for the pre 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 year of transaction, road name, main floor area, storey, land area, roof material, floor material, wall material, ceiling material, maintenance inside, maintenance outside, architectural functionalistic, historical styles, ensemble, authenticity, multifunction and position.

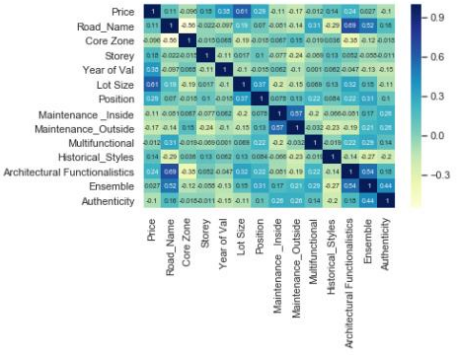
Table 1shows the filtering process of the original set of data from 2004 to 2018 in which 137 observations (pre war shop house) remained for this study. The data were examined for completeness and usefulness to develop the training machine learning algorithms.

Table 1: A record of data cleaning process

No	Data	Number of records left
1.	Original data from 2004 to 2018 for pre war 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	137

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 & Tsolacos, 2010). Figure 2 shows the correlation indices of the original variables and python heatmap plot of all variables in order to indicate the multicollinearity among them. The results reveal that there are no variables that have collinearity index above 0.8. The darkest color represents the strongest relationship among the two variables.




```
Out[123]: Price                          1.000000
          TransformedPrice               0.845756
          Lot Size                       0.613839
          Year of Val                    0.378055
          Position                       0.285980
          Architectural Functionalistics 0.238780
          Storey                         0.180882
          Historical_Styles              0.136962
          Road_Name                      0.110198
          Ensemble                       0.026798
          Multifunctional               -0.011578
          Core Zone                     -0.096344
          Authenticity                  -0.104773
          Maintenance _Inside           -0.110320
          Maintenance_Outside           -0.171661
          Id                            -0.818019
          Name: Price, dtype: float64
```

4.0 RESULTS AND DISCUSSION

This section compares the predictive performance of heritage price prediction model based on machine learning algorithms which are random forest regressor, decision tree regressor, lasso, ridge, linear regression. 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 2. For this task, we used the Pyhton software. Based on the result of pre war shop houses, random forest regressor is the best algorithm in predicting the price of prewar shophouses with the least RMSE, and highest R squared values.

Table 2: Performance indicators for five algorithms

Algorithms	Random Forest Regressor	Decision Tree Regressor	Lasso	Ridge	Linear Regression
R squared	**0.997**	0.996	0.951	0.927	0.925
RMSE	**0.034651**	0.040408	0.147119	0.179620	0.182312

a) 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,
b) R squared 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.

5.0 CONCLUSIONS

This study has presented the empirical results based on 137 pre war shop houses in North-East, Penang Island, Malaysia using machine learning techniques. Several machine learning algorithms have been used to develop prediction models for pre war shophouses in North-East Penang Island. Five different supervised machine learning algorithms used are random forest regressor, decision tree regressor, lasso, ridge, linear regression. The findings show that the random forest regressor model has produced the best predictions of pre-war shophouses in North East, Penang Island, Malaysia.

In conclusion, this paper has revealed the potential of machine learning in predicting the price of heritage property by highlighting the superiority of random forest regressor algorithms.

Our next endeavour will be to compare between machine learning algorithms, multiple regression analysis, rank transformation regression and contingent valuation method.

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

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

Horino, H., Nonaka, H., Carreón, E. C. A., & Hiraoka, T. (2017). Development of an entropy-based feature selection method and analysis of online reviews on real estate. In 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) (pp. 2351–2355). IEEE.

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

Kilibarda, M. (2018). Estimating the Performance of Random Forest versus Multiple Regression for Predicting Prices of the Apartments. Journal of Geo-Information, 1–16. https://doi.org/10.3390/ijgi7050168

Muller, A. C., & Guido, S. (2017). Introduction to machine learning with scikit-learn. Kaggle's Blog.

Ng, A., & Deisenroth, M. (2015). Machine learning for a London housing price prediction mobile application. In Imperial College London.

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) (pp. 35–42). IEEE.

Sayce, S. (2009). Valuing Heritage Assets. (RICS & H. M. Treasury, Eds.), Examining the Case for the Valuationof Heritage Assets. Kingston University London.

Varma, A., Sarma, A., Doshi, S., & Nair, R. (2018). House Price Prediction Using Machine Learning And Neural Networks. Journal of Inventive Communication and Computational Technologies (ICICCT), 1936–1939.

Yung, E. H. K., Yu, P. L. H., & Chan, E. H. W. (2013). Economic Valuation of Historic Properties: Review and Recent Developments. Property Management, 31(4), 335–358. https://doi.org/10.1108/PM-01-2013-0005

Pejabat Perpustakaan
Librarian Office

Universiti Teknologi MARA
Cawangan Perak
Kampus Seri Iskandar
32610 Bandar Baru Seri Iskandar,
Perak Darul Ridzuan, MALAYSIA
Tel: (+605) 374 2093/2453 Faks: (+605) 374 2299

UNIVERSITI TEKNOLOGI MARA

Surat kami : 700-KPK (PRP.UP.1/20/1)
Tarikh : 20 Januari 2023

UNIVERSITI TEKNOLOGI MARA PERAK
TERIMA
25 JAN 2023
Tindakan
Universiti Teknologi MARA Perak
PEJABAT REKTOR

Prof. Madya Dr. Nur Hisham Ibrahim
Rektor
Universiti Teknologi MARA
Cawangan Perak

Tuan,

PERMOHONAN KELULUSAN MEMUAT NAIK PENERBITAN UiTM CAWANGAN PERAK MELALUI REPOSITORI INSTITUSI UiTM (IR)

Perkara di atas adalah dirujuk.

2. Adalah dimaklumkan bahawa pihak kami ingin memohon kelulusan tuan untuk mengimbas (*digitize*) dan memuat naik semua jenis penerbitan di bawah UiTM Cawangan Perak melalui Repositori Institusi UiTM, PTAR.

3. Tujuan permohonan ini adalah bagi membolehkan akses yang lebih meluas oleh pengguna perpustakaan terhadap semua maklumat yang terkandung di dalam penerbitan melalui laman Web PTAR UiTM Cawangan Perak.

Kelulusan daripada pihak tuan dalam perkara ini amat dihargai.

Sekian, terima kasih.

"BERKHIDMAT UNTUK NEGARA"

Saya yang menjalankan amanah,

SITI BASRIYAH SHAIK BAHARUDIN
Timbalan Ketua Pustakawan

nar

Setuju.

27.1.2023

PROF. MADYA DR. NUR HISHAM IBRAHIM
REKTOR
UNIVERSITI TEKNOLOGI MARA
CAWANGAN PERAK
KAMPUS SERI ISKANDAR