

Machine learning building price prediction with green building determinant

Thuraiya Mohd¹, Syafiqah Jamil², Suraya Masrom³

¹Faculty of Architecture, Planning and Surveying, Universiti Teknologi MARA, Perak Branch Malaysia

²Faculty of Architecture, Planning and Surveying, Universiti Teknologi MARA, Shah Alam Selangor, Malaysia

³Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch Tapah Campus, Malaysia

Article Info

Article history:

Received Feb 6, 2020

Revised Apr 16, 2020

Accepted May 4, 2020

Keywords:

Algorithm
Green building
Machine learning
Python
Prediction
Skirt-learn

ABSTRACT

In the era of Industrial 4.0, many urgent issues in the industries can be effectively solved with artificial intelligence techniques, including machine learning. Designing an effective machine learning model for prediction and classification problems is an ongoing endeavor. Besides that, time and expertise are important factors that are needed to tailor the model to a specific issue, such as the green building housing issue. Green building is known as a potential approach to increase the efficiency of the building. To the best of our knowledge, there is still no implementation of machine learning model on GB valuation factors for building price prediction compared to conventional building development. This paper provides a report of an empirical study that model building price prediction based on green building and other common determinants. The experiments used five common machine learning algorithms namely Linear Regression, Decision Tree, Random Forest, Ridge and Lasso tested on a set of real building datasets that covered Kuala Lumpur District, Malaysia. The result showed that the Random Forest algorithm outperforms the other four algorithms on the tested dataset and the green building determinant has contributed some promising effects to the model.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Thuraiya Mohd,
Faculty of Architecture, Planning and Surveying,
Universiti Teknologi MARA, Perak Branch Malaysia.
Email: Thura231@uitm.edu.my

1. INTRODUCTION

The advancement of 4.0 industrial revolution along with the rising interest of big data technology has catalysed the importance and rapid-development of data science field. Data science has played an important role in many industries including medical diagnosis [1], cheminformatics and bioinformatics [2], stock market analysis [3], detecting credit card fraud [4] and many more. The reasons behind the increasing interest include the availability of data, a variety of open-source machine learning tools and powerful computing resources.

The Green Building (GB) can be defined as an approach of increasing efficiency of the building and sites by using energy, water and natural materials. It can also reduce the impact on human, environment and health by improving system operation, maintenance, design, construction and transfer of the complete building life cycle [5-6]. Green Building Index (GBI) is a distinguished Malaysian industry that recognizes green rating tool, in which it helps to determine the types of building categorized as Platinum, Gold and Silver.

In the Malaysian real estate industry, GB is still in its infancy, where the valuation is not integrated into standard property valuation [7]. The valuation standard only provides the valuation of the property and buildings, which may not have sufficient definition to include GB development [8-9]. It creates some difficulties for the valuers to assimilate the conventional method of valuation to indicate and predict the price of GB accurately [10]. The problem arises as researchers see another problem related to real estate transaction data. It is stated that valuers often face difficulties to predict property prices over the time [11], especially when a matter related to limitation of data evidence transaction on GB valuation because GB development is relatively new in Malaysia and comparatively new in the real estate industry [12-13]. The valuation of non-GBs often depends on leasing or sales transaction data from several properties provided by JPPH and the data is unlimited. It is important to realise that valuers face various challenges because of their heavy dependency on market data. Lack of data means lack of support for the valuable contributions of green attributes, which is supposed to be the factor influencing the GB price. Indeed, the real estate market is exposed to many price fluctuations due to existing correlations with many variables and some of which are beyond our control or perhaps unknown [14].

In light of this situation, Machine learning (ML) model has emerged as a very promising approach in resolving the issue and it is proven to be effective in different kinds of prediction and classification problem [15-17]. ML model has different kinds of algorithms and techniques to be selected for developing a good predictor model. These are beneficial to resolve dataset problems such as imbalance and insufficient data like the limitation of sale data evidence transaction of GB valuation. However, the accuracy of the results produced by the ML prediction model is highly dependent on many factors including the algorithms hyper-parameters tuning and different group of features selection. Thus, this paper is written with the aim to report the design and implementation of machine learning model based on auto hyper-parameters tuning and different groups of feature selection.

The contribution of this paper is two-fold. Firstly, it introduces the design and implementation of machine learning model with auto hyper-parameter tuning. In the methodology part, this paper provides the technique of auto hyper-parameter tuning by using best estimator function provided by *Python Scikit-Learn* library. Secondly, it presents how GB determinant affects the machine learning performance in predicting the price of building based on real dataset of Kuala Lumpur district in Malaysia

The structure of this paper is as follows. Section II focuses on the background of the study related to the ML in real prediction of real estate and ML algorithms. Section III describes the research methodology followed by the discussion of the result in section IV. The concluding remark is written in the last section.

2. BACKGROUND OF THE STUDY

2.1. Machine learning for real estate prediction

Accurate evaluation of property price is crucial for real estate, the stock market, tax sector, the economy and the power of purchasers [18]. The conventional method is limited to the scope of current systems data that needs to be taken into account. Normally, predicting the price of property is often done through basic comparative market analysis as well as similar real estate in the same area to provide an approximate price for a particular property [19]. But in GB context, the other factors that can contribute or give positive impact or added values to the GB price should also be considered to produce an accurate result in the price and to reflect the current market value [20]. This will only happen if the valuer considers the historical factors in predicting the price of the GB. ML is seen to have the potential in considering those factors and problems [14].

The common ML modelling techniques that are already being implemented in real estate problems are Linear Regression [21-23], Decision Tree [24-27], Random Forest [21, 28-29], Ridge Regression [30] and Lasso Regression [24, 31]. The function of all these algorithms is to predict the real estate dataset and the researchers test all these algorithms in order to predict the green building prices.

2.2. Machine learning algorithm

There are five (5) ML algorithms that are used in this study namely Linear Regression, Decision Tree, Random Forest, Ridge and Lasso algorithms.

Linear Regression (LR) is one of the most well-understood and well-known algorithms in ML and statistics. It is also a predictive model that mainly concerns in minimising the error and to ensure or to make the most accurate and possible prediction in explaining the dataset ability. The representation of LR algorithm is an equation that explains and describes a line which ensures the best fits of the relationship between the output variables (y) and input variables (x), by finding the exact weighting for the input variable that is called coefficient (B) [32]. The formula in (1) representing the Linear Regression algorithm.

$$Y = B_0 + B_1 * x \quad (1)$$

In this formula, Y is the dependent variable (DV) by the given input (x) which is the independent variable (IV). The main goal of the Linear Regression algorithm is to find the value for the coefficients B_0 and B_1 [21-22, 25]. Due to the simplicity of algorithm, Linear Regression has been commonly used in real estate prediction problem [13-15].

Decision Tree (DT) is another common model used to solve regression and classification problem [33]. The algorithm produces a tree structure that includes a root node and branches. Each internal node stands for a test on an attribute, each branch denotes the outcome of a test, which is called a decision node and each leaf node holds a class label which is called a terminal node. The topmost of the node in the tree is called a root node [33-34] as presented in Figure 1.

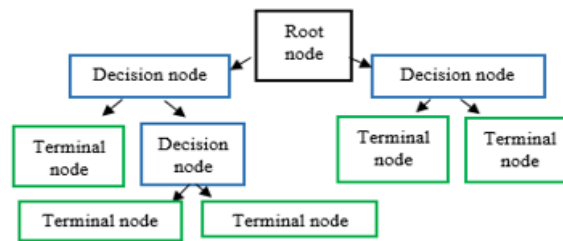


Figure 1. Decision tree structure

However, previous research showed the designs which indicate that the DT algorithm can provide a higher accuracy to dataset, compared to the other algorithm like Lasso [24]. DT has no problems in approximating the linear relationships based on Independent Variable and Dependent Variable factors [25-26]. It is good to perform the algorithm when it comes to prediction.

The Random Forest (RF) is an advanced tree structures from the DT, [35-38]. It is a type of ensemble ML model called Bootstrap, Bagging or Aggregation. The bootstrap is a powerful statistical method for estimating a quantity from a data sample such as the mean. RF model will take a lot of data samples, calculate the mean, then average all of the mean values to give a better estimation result of the true mean value [39]. Several research have demonstrated that RF mostly outperforms many other algorithm in dealing with problem related to property price [21, 28-29].

The Ridge algorithm is one of ML models that is used for analysing multiple regression dataset that suffers from multicollinearity. Multicollinearity is also called as collinearity that refers to a position in which two or more informative variables in a multiple regression are highly related. Even though, Ridge Regression algorithm is added in that problem, a degree of bias to the regression can still be estimated. Ridge Regression is a model that enforces the coefficient to be lower but it does not enforce them to be zero, as it will not get rid of irrelevant feature but rather minimising their impact on the training model [40]. To avoid overfitting, Ridge Regression algorithm performs L2 regularisation stated in the formula. Meanwhile, Lasso algorithm uses L1 regularisation [41]. Equation (2) denotes Ridge algorithm.

$$Y = XB + e \quad (2)$$

In this formula, Y denotes for DV, X as IV and B represents the regression of coefficient to be predicted [40]. The e represents the residual errors. There are some research which prove that Ridge Regression can be less performed compared to Linear Regression although the Ridge Regression is designed to handle multicollinearity in modelling house price [30]. In the other study on house price prediction, Lasso Regression has outperformed Ridge algorithm in handling multicollinearity. Furthermore, in real estate value prediction using multiple algorithm, Lasso regression algorithm seems to overfit their model dataset by using Ridge Regression algorithms [42].

Lasso regression algorithm stands for Least Absolute Selection and Shrinkage Operator and it can perform both tasks which are feature selection and regularisation. The only difference of Lasso algorithm from Ridge Regression algorithm is that the regularisation term is in absolute value. It is set to restrain the sum of the absolute values of the model parameters where the sum must be less than a fixed value [43-44]. Besides that, Lasso algorithm is being applied in a shrinking (regularisation) process where it penalizes the

coefficients of the regression variables shrinking some of them to zero if they are not relevant. Indeed, this process is being applied to minimise the prediction error.

Research in [24] has demonstrated the potential of Lasso algorithm to produce higher accuracy than Linear regression and decision tree within the scope of study. The algorithm was employed in predicting the house price in Ames, Iowa in United State using real estate data from 2016 to 2020 and it was found that Lasso algorithm outperformed Ridge algorithm in this case [30]. The researchers also mentioned that Lasso is very useful for features selection and to eliminate any useless features.

3. METHODOLOGY

3.1. Dataset

The dataset is a collection of housing prices in 2018 with determinants that includes GB. As this paper uses machine learning prediction, these variables are called features. Table 1 shows the set of features to develop the machine learning prediction model. This study uses 18 features as independent variables (IV) for predicting the Transaction Price as dependent variable (DV).

Table 1. Feature in the dataset

Features	Description
Transaction Price	Dispose price/sqf (RM)
Date of Transaction	Building Transaction/Months
Lot Size	Lot Size
MFA	Main Floor Area
Tenure	Freehold/Leasehold
Type of Property	Residential/Commercial
No of bedroom	Number of bedrooms
Level Property	Level Property Unit
Floor	Building Floor
Building Facade	City/Park/Lake/Klcc
Age of Building	Age of Building
Distance	Distance to Central Business District
Accessibility	Ease of accessibility
Mukim	Mukim
Certificate	Green Certificate/Non-Green Building
Density	Population Density
Security	Security of Building
Infrastructure	Infrastructure Development

3.2. Feature selection

The following Figure 2 shows the Pearson correlation between all features to the DV running with Python codes.

All the IVs have a very weak correlation to the Transaction Price. GB variable has the highest correlation among the features but 0.25 is considered weak. However, even with a very weak correlation, it was anticipated in the study that to some degree they still contribute impressive information to the model.

```
In [169]: correlations = training.corr()
correlations = correlations["Transaction Price"].sort_values(ascending=False)
features = correlations.index[1:6]
correlations

Out[169]: Transaction Price      1.000000
GREEN CERTIFICATE      0.253264
NO. OF BEDROOM         0.146792
FLOOR                  0.103781
Type of property       0.093225
LEVEL PROPERTY UNIT    0.083747
MFA                    0.036486
M UKIM                 0.035657
luas lot               0.029245
Date of Transaction    0.015224
Tenure                 0.009775
ACCESSIBILITY          0.003045
SECURITY OF BUILDING  -0.009663
Distance               -0.011069
Age of building        -0.030878
BUILDING FAÇADE        -0.115872
POPULATION DENSITY    NaN
INFRASTRUCTURE DEVELOPMENT NaN
Name: Transaction Price, dtype: float64
```

Figure 2. Pearson correlation between DV and Ivs (features)

There are several approaches in selecting the features for machine learning model. They can be divided according to the features correlation level or based on the feature's types or purposes. In this study, features were divided into three types namely without GB, GB and GB with other features.

3.3. Machine learning algorithms with auto-hyper parameter tuning

The five algorithms explained in part 2.2 namely Random Forest Regressor, Decision Tree Regressor, Ridge, Lasso and Linear Regression were used in this study. Prior to the prediction results prediction, auto hyper-parameter tuning was implemented first based on the training dataset by calling *best_estimator* method in the *Python Scikit-Learn* library. The method uses grid search optimization of hyper-parameter tuning on the given machine learning algorithm. This is the easiest and shortest time ways for inexperienced data scientist to get the suggestions of parameters configuration for the algorithms.

The steps of implementing the auto hyper-parameters are as follow:

1. Call the regressor algorithm.
2. Create dictionary and define initial parameters for the algorithm with the corresponding set of values.
3. Call the grid search method by passing the created dictionary.
4. Do preliminary training for the algorithm with the grid search instance and get the parameters estimator.
5. Set the algorithm with the suggested parameters and conduct another fitness with the suggested parameters.
6. Perform another training with the suggested parameters.
7. Validate the prediction value produced by the algorithm and get the score.

3.4. Experiment configuration

In this study, the training and validation datasets were divided into the ratio 80:20 respectively. Python 3.6 Jupyter Notebook platform with Intel i7 7th Generation processor on 16 GB RAM were used. Each machine learning model with each algorithm was set to employ 80:20 percent ratios between the training and validation separation. Each model was run for five times of experiments and the average results of metrics were calculated for comparison. The metrics to present the performances of machine learning algorithms are R squared (R^2) and root mean squared error (RMSE). The R^2 can explain how well the selected features in predicting the dependent variable while RMSE represents the sample standard deviation the difference between the predicted and real values. The range of values for R^2 is between 0.1 with higher is better. Meanwhile, RMSE with lower value shows lower errors or differences in the prediction results.

4. RESULT

The results are presented in different tables according to the three groups of features selection namely without GB, GB only and GB with other features. The average results from the five times experiments of each machine learning model were calculated and recorded. The results of model without GB features selection is presented in Table 2.

Table 2. Result of machine learning algorithms without GB

Algorithm	R^2	RMSE
Random Forest Regressor	0.693	0.027
Decision Tree Regressor	0.180	0.053
Ridge	0.035	0.048
Linear Regression	0.051	0.048
Lasso	0.000	0.045

Without GB determinant, only the Random Forest Regressor could produce an acceptable result. The algorithm had the lowest RMSE (0.027) and the highest coefficient of determination presented by the R^2 (0.69). The mean of R^2 from other algorithms appeared to be very weak but the errors distanced of each algorithm is considered promising. The following Table 3 presents the mean of R^2 and RMSE for the tested algorithms with GB determinant only.

Similarly, Random Forest regressor outperformed other algorithms but the values for RMSE and R^2 were not as good as the value in Table 2. The performances of Random model regressor dropped when only dependent on the GB determinant. However, not much different could be seen on the other algorithms. Lastly, Table 4 lists the results of each algorithm when tested with all determinants that combined GB and others.

Table 3. Results of machine learning algorithm with GB only

Algorithm	R ²	RMSE
Random Forest Regressor	0.145	0.046
Decision Tree Regressor	0.00	0.0492
Ridge	0.028	0.0485
Linear Regression	0.046	0.048
Lasso	0.000	0.0492

Table 4. Result of machine learning algorithms with GB and other feautres

Algorithm	R ²	RMSE
Random Forest Regressor	0.647	0.0292
Decision Tree Regressor	0.330	0.040
Ridge	0.060	0.0477
Linear Regression	0.095	0.048
Lasso	0.000	0.047

Combining GB with other features in the models does not really show a significant improvement to each of the tested algorithms. Slightly better performance can be seen on the Decision Tree regressor for the R².

5. CONCLUSION

Within the scope of this study, it can be concluded that GB determinant has not contributed much to the performance of machine learning models even though its correlation to the building price is higher than the other determinants. Moreover, the worst results of all algorithms produced by the model with single GB determinant. Among the five selected algorithms, only Random Forest regressor shows a consistent performance with all the group of features selection. Therefore, Random Forest regressor can be further enhanced in future research for the same case of building price prediction.

ACKNOWLEDGEMENTS

The researcher would like to thank Ministry of Education MALAYSIA and UNIVERSITI TEKNOLOGI MARA for the funding granted to this project under FRGS grant FRGS/1/2018/SS08/UITM/02/5.

REFERENCES

- [1] A. Key, T. Sch, and J. A. Med, "Artificial Intelligence & Medical Diagnosis," *Sch. J. Appl. Med. Sci. (SJAMS) Artif. Intell. Med. Diagnosis*, no. December, pp. 4982-4985, 2019.
- [2] J. B. O. Mitchell B.O., "Machine Learning Methods in Chemoinformatics," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 4, no. 5, pp. 468-481, 2014.
- [3] K. H. Sadia, A. Sharma, A. Paul, S. Padhi, and S. Sanyal, "Stock Market Prediction Using Machine Learning Algorithms," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 4, pp. 25-31, 2019.
- [4] O. S. Yee, S. Sagadevan, and N. H. A. H. Malim, "Credit Card Fraud Detection Using Machine Learning as Data Mining Technique," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 1-4, pp. 23-27, 2018.
- [5] N. Ismail, M. N. Rahmat, and S. Y. Said, "Proceedings of the Colloquium on Administrative Science and Technology," *New Sustain. Heritage-led urban Regen. mode*, no. October, pp. 2-7, 2015.
- [6] M. R. Radwan, A. E. B. Kashyout, H. G. Elshimy, and S. F. Ashour, *Green Building as Concept of Sustainability Sustainable Strategy to Design Office Buidling*, no. April vol.2. 2015.
- [7] M. O. Ajibola and O. O. Awodiran, "Effects of Infrastructure on Property Values in Unity Estate, Lagos, Nigeria," *Int. J. Econ. Manag. Soc. Sci.*, vol. 2, no. May, pp. 195-201, 2013.
- [8] S. M. Ab.Aziz, S.S, Sipan, i, "The Effect of Green Envelope Components on Green Building Value," *Prop. Manag. J.*, vol. 35, no. 2, pp. 181-201, 2017.
- [9] L. Abdullah, W. N. A. W. A. Rasid, and T. Mohd, "The Role of Valuer in Sustainable Valuation: A Review," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 8, no. 2, pp. 740-749, 2018.
- [10] R. A. Majid, "The Impact of Green Features on Property Valuation Procedure," *Natl. Inst. Valuat. (INSPEN), Putrajaya*, no. April 2014, pp. 4-15, 2016.
- [11] Ž. Sabina and M. Trojanek, "Current Problem of Valuation and Real Estate Management BY Value," *Croat. Inf. Technol. Soc. GIS Forum, Croat. Univ. Warm. Maz. Olsztyn, Pol. Univ. Silesia, Pol.*, no. January, pp. 1-20, 2014.
- [12] G. W. Austin, "Sustainability and Income- Producing Property Valuation: North American Status and Recommended Procedures," *Am. Valuat.*, vol. Vol.4, pp. 79-81, 2009.

- [13] M. Isa, M. Mohamed, G. Megat, A. Rahman, and I. Sipan, "Factors Affecting Green Office Building Investment in Malaysia Factors affecting Green Office Building Investment in Malaysia," *Procedia - Soc. Behav. Sci.*, vol. 105, no. January 2014, pp. 138-148, 2013.
- [14] A. Baldominos, I. Blanco, A. Moreno, R. Iturrarte, Ó. Bernárdez, and C. Afonso, "Identifying Real Estate Opportunities Using Machine Learning," *Appl. Sci.*, vol. 8, no. 11, pp. 2-24, 2018.
- [15] D. K. Choubey, P. Kumar, S. Tripathi, and S. Kumar, "Performance Evaluation of Classification Methods with PCA and PSO for Diabetes," *Netw. Model. Anal. Heal. Informatics Bioinforma.*, vol. 9, no. 1, pp. 5-9, 2020.
- [16] J. Ayoub, D. Lotfi, M. El Marraki, and A. Hammouch, "Accurate Link Prediction Method Based on Path Length Between a Pair of Unlinked Nodes and their Degree," *Soc. Netw. Anal. Min.*, vol. 10, no. 1, 2020.
- [17] T. Cura, "Use of Support Vector Machines with a Parallel Local Search Algorithm for Data Classification and Feature Selection," *Expert Syst. Appl.*, vol. 145, pp. 113-133, 2020.
- [18] E. Pagourtzi, V. Assimakopoulos, T. Hatzichristos, and N. French, "Real Estate Appraisal: A Review of Valuation Methods," *J. Prop. Invest. Financ.*, vol. 21, no. 4, pp. 383-401, 2003.
- [19] M. Kummerow, "Theory for Real Estate Valuation : An Alternative Way to Teach Real Estate Price Estimation Methods," *L. Econ. Valuation, Curtin Univ. Perth*, pp. 1-29, 2003.
- [20] L. Abdullah and T. Mohd, "Green Building Valuation ; From a Valuers ' Perspective," *Adv. Civ. Eng. Sci. Technol.*, no. October, pp. 3-9, 2018.
- [21] S. Borde, A. Rane, G. Shende, and S. Shetty, "Real Estate Investment Advising Using Machine Learning," *Int. Res. J. Eng. Technol.*, vol. 4, no. March, pp. 1821-1825, 2017.
- [22] T. Dimopoulos, H. Tyralis, N. P. Bakas, and D. Hadjimitsis, "Accuracy Measurement of Random Forests and Linear Regression for Mass Appraisal Models that Estimate the Prices of Residential Apartments in Nicosia, Cyprus," *Creat. Commons Attrib. 4.0 Licens.*, no. November, pp. 1-6, 2018.
- [23] M. Van Wezel and R. Potharst, "Boosting the Accuracy of Hedonic Pricing Models," *Econom. Inst. Rep. EI 2005-50*, no. May, pp. 1-18, 2005.
- [24] N. Shinde and K. Gawande, "Valuation of House Prices Using Predictive Techniques," *Int. J. Adv. Electron. Comput. Sci. ISSN2393-2835*, vol. Volume-5, no. 6, pp. 34-40, 2018.
- [25] M. Van Wezel and R. Potharst, "Boosting the Accuracy of Hedonic Pricing Models," *Econom. Inst. Rep. EI 2005-50*, vol. 2, no. December, pp. 1-18, 2005.
- [26] Y. Ma, Z. Zhang, A. Ihler, and B. Pan, "Estimating Warehouse Rental Price using Machine Learning Techniques," *Int. J. Comput. Commun. Control ISSN*, vol. 13, no. April, pp. 235-250, 2018.
- [27] A. Baldominos, I. Blanco, A. J. Moreno, R. Iturrarte, Ó. Bernárdez, and C. Afonso, "Identifying Real Estate Opportunities Using Machine Learning," *Appl. Sci. Mdpi J.*, pp. 2-13, 2018.
- [28] W. Lee, N. Kim, Y. Choi, Y. S. Kim, and B. Lee, "Machine Learning based Prediction of the Value of Buildings," *KSII Trans. Internet Inf. Syst.*, vol. 12, no. 8, pp. 3966-3991, 2018.
- [29] M. A. Valle, "Property Valuation using Machine Learning Algorithms : A Study in a Metropolitan-Area of Chile," *Int. Conf. Model. Simul. Santiago Chile*, no. May 2017, pp. 1-133, 2016.
- [30] S. J. Xin and K. Khalid, "Modelling House Price Using Ridge Regression and Lasso Regression," *Int. J. Eng. Technol.*, vol. 7, pp. 498-501, 2018.
- [31] S. Lu, Z. Li, X. Qin, X. Yang, R. Siow, and M. Goh, "A Hybrid Regression Technique for House Prices Prediction," *Conf. Pap. Inst. High Perform. Comput.*, no. August 2018, pp. 1-5, 2017.
- [32] A. Schneider, G. Hommel, and M. Blettner, "Lineare Regressions Analysis," *Dtsch. Arztebl.*, vol. 107, no. 44, pp. 776-782, 2010.
- [33] W. Loh, "Classification and regression trees," *WIREs Data Min. Knowl. Discov.*, vol. 1, no. February, pp. 1-14, 2011.
- [34] H. Sharma and S. Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining," *Int. J. Sci. Res.*, vol. 5, no. 4, pp. 2094-2097, 2016.
- [35] J. Izenman, *Modern Multivariate Statistical Techniques*, vol. 102. 2006.
- [36] J. Franklin, "The Elements of Statistical Learning: Data Mining, Inference and Prediction," *Math. Intell.*, vol. 27, no. 2, pp. 83-85, 2005.
- [37] C. Strobl, J. Malley, and G. Tutz, "An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests," *Psychol. Methods*, vol. 14, no. 4, pp. 323-348, 2009.
- [38] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random Forests," in *Chapter from book Ensemble Machine Learning pp.157-176*, no. February 2014, 2011, p. 157.
- [39] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random Forests and Decision Trees," *Int. J. Comput. Sci. Issues*, vol. 9, no. 5, pp. 272-278, 2012.
- [40] E. Fox and C. Guestrin, "Ridge Regression," in *NCSS Statistical Software*, 2015, pp. 1-21.
- [41] L. C. Bergersen, "Guiding the Lasso : Regression in High Dimensions," *Dep. Math. Univ. Oslo, Present. degree Philos. Dr.*, pp. 6-10, 2013.
- [42] R. Manjula, S. Jain, S. Srivastava, and P. Rajiv Kher, "Real Estate Value Prediction using Multivariate Regression Models," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 263, no. 4, pp. 2-8, 2017.
- [43] J. M. Pereira, M. Basto, and A. F. da Silva, "The Logistic Lasso and Ridge Regression in Predicting Corporate Failure," *Procedia Econ. Financ.*, vol. 39, no. November 2015, pp. 634-641, 2016.
- [44] V. Fonti, "Feature Selection Using LASSO," *VU Amsterdam*, pp. 1-26, 2017.

BIOGRAPHIES OF AUTHORS

Associate Professor in Universiti Teknologi MARA. Graduated with PhD in Real Estate in 2012 and MSc in Land Development and Administration in 2003 from Universiti Teknologi Malaysia. Received Bachelor's degree in Estate Management from Universiti Teknologi MARA in 2001. Prior to joining Universiti Teknologi MARA, she served with Ismail & Co as a Valuation Executive for 2 years. Lecturing experience in UiTM, Perak Branch for 18 years in the Department of Estate Management. Lecturing experience includes teaching core courses of Valuation and Property Development to undergraduate students, Built Environment Theory to PhD students and Economics of Green Architecture to Masters students. She was awarded an Excellent University Community Transformation Centre (UCTC) Award 2015 by the Ministry of Education (MOE), MALAYSIA. Received few grants (as leader) from the MOE and Ministry of Finance, MALAYSIA in the areas of property development and disaster management. Current research interests are focused in the areas of green development, housing development, machine learning and disaster management. She was appointed as Coordinator of Academic Development and Management UiTM Perak Branch. She achieved Professional Qualifications as Surveyor (Sr) and is a professional member of the Board of Valuers, Appraisers, Estate Agents and Property Managers (BOVEAP)



Syafiqah Jamil graduated with a Diploma in Real Estate Management in 2017. She began her first career in the industry when she worked as a valuation Executive on Nilai Harta Consultants in May 2016, and Valuation Executive at Md Nor and Partner companies in December 2016, Where her work focused on real estate finance loans, auction cases, refinancing and remortgage. In addition, he also has experience as a Real Estate Agent with MRCB developers as a part-time job. She then pursued a Bachelor's degree of Real Management in March 2017 at UiTM Campus, Shah Alam and received a Bachelor's degree of Real Estate Management in 2019. Currently, she is studying at UiTM, Shah Alam Campus, (Built Environment) under a research grant, where current research is focused on Green Building and Machine Learning.



Suraya received her bachelor's degree in Computer Science, majoring in software engineering, from UTM. She later pursued her master's degree in Computer Science at UPM. She first started her career in industry when she was employed as an Associate Network Engineer by Ramgate Systems Sdn. Bhd in June 1996. She started her career as a full-time lecturer with UTM after receiving her master's degree. In three years of service at UTM she managed to complete two research projects funded by the university. She was offered a position in University Teknologi MARA (UiTM), Seri Iskandar, Perak, Malaysia, in 2004, which she gladly accepted and has been lecturing Computer Science subjects there until presently. In 2015, she received her PhD in Information Technology and Quantitative Sciences. In UiTM, she has so far managed to complete more than fifteen research projects and is currently active with three research grants.