

House Price Prediction using a Machine Learning Model: A Survey of Literature

Nor Hamizah Zulkifley

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor
Email: mizalee10@gmail.com

Shuzlina Abdul Rahman

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor
Email: shuzlina@fskm.uitm.edu.my

Nor Hasbiah Ubaidullah

Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris, 35900 Tg. Malim Perak.
Email: hasbiah@fskik.upsi.edu.my

Ismail Ibrahim

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor
Email: ismailibrahim@uitm.edu.my

Received: 15 July 2020; Accepted: 22 October 2020; Published: 08 December 2020

Abstract: Data mining is now commonly applied in the real estate market. Data mining's ability to extract relevant knowledge from raw data makes it very useful to predict house prices, key housing attributes, and many more. Research has stated that the fluctuations in house prices are often a concern for house owners and the real estate market. A survey of literature is carried out to analyze the relevant attributes and the most efficient models to forecast the house prices. The findings of this analysis verified the use of the Artificial Neural Network, Support Vector Regression and XGBoost as the most efficient models compared to others. Moreover, our findings also suggest that locational attributes and structural attributes are prominent factors in predicting house prices. This study will be of tremendous benefit, especially to housing developers and researchers, to ascertain the most significant attributes to determine house prices and to acknowledge the best machine learning model to be used to conduct a study in this field.

Index Terms: House Price Prediction, Machine Learning Model, Support Vector Regression, Artificial Neural Network, XGBoost

1. Introduction

House is one of human life's most essential needs, along with other fundamental needs such as food, water, and much more. Demand for houses grew rapidly over the years as people's living standards improved. While there are people who make their house as an investment and property, yet most people around the world are buying a house as their shelter or as their livelihood.

According to [1], housing markets have a positive impact on a country's currency, which is an important national economy scale. Homeowners will purchase goods such as furniture and household equipment for their home, and homebuilders or contractors will purchase raw material to build houses to satisfy house demand, which is an indication of the economic wave effect created by the new house supply. Besides that, consumers have capital to make a large investment, and the construction industry is in good condition can be seen through a country's high level of house supply.

According to [2], numerous international organizations and human rights have emphasized house importance. House is profoundly rooted in the economic, financial, and political structure of each country. Nevertheless, [3] reported that the fluctuation of house prices has always been an issue for house owners, buildings and real estate, besides [4] stated that house has become unaffordable as there is substantial price growth in several countries in the

housing sector. Residents' quality of life as well as national economy depends on the potential house price increase. Ultimately, this issue will affect investors who are making their house as an investment.

An increase in house demand occurs each year, indirectly causing house price increases every year. The problem arises when there are numerous variables such as location and property demand that may influence the house price, thus most stakeholders including buyers and developers, house builders and the real estate industry would like to know the exact attributes or the accurate factors influencing the house price to help investors make decisions and help house builders set the house price.

House price prediction can be done by using a multiple prediction models (Machine Learning Model) such as support vector regression, artificial neural network, and more. There are many benefits that home buyers, property investors, and house builders can reap from the house-price model. This model will provide a lot of information and knowledge to home buyers, property investors and house builders, such as the valuation of house prices in the present market, which will help them determine house prices. Meanwhile, this model can help potential buyers decide the characteristics of a house they want according to their budget [5]. Previous studies focused on analyzing the attributes that affect house price and predicting house price based on the model of machine learning separately. However, this article combines such a both predicting house price and attributes together.

In this article, literature review focuses on predicting house price based on the model of machine learning as well as analyzing attributes primarily used in previous study that affect house price. This paper was arranged as follows: the first section summarizing overall of this study. Second section described the common attributes used in prediction of house price around the world. It was followed by a brief discussion of machine learning model used in previous study to predict house price. For the next section, the comprehensive effects of the current house price prediction model are addressed. Ultimately, section 5 and section 6 respectively provide the description and conclusion of this comprehensive literature analysis.

2. Attributes

House price prediction can be divided into two categories, first by focusing on house characteristics, and secondly by focusing on the model used in house price prediction. Many researchers have produced a house price prediction model, including [1, 3, 6–8].

A research undertaken by [9] analyses the existing housing price in Jakarta, Indonesia using the conceptual model and questionnaires. Based on the results, the attributes or factors affecting the house price differ for each house construction in Jakarta, therefore accepting the validity of this analysis as the main purpose of this research is to classify the factor or attributes affecting the house price. Various considerations influence the price of a house. According to [10], the factors influencing house prices can be classified into three categories: location, structural and neighborhood condition.

A. Locational

Location is considered to be the most significant feature of house price determination [6, 9–11]. [12] in his study also observed the significant of location attributes in deciding house price. The location of the property was classified in a fixed locational attribute. All of these studies point to the close association between locational attributes such as distance from the closest shopping center, or position offering views of hills or shore, and house price variations.

B. Structural

Another significant feature influencing the house price is structural structure or some research has listed it as physical attributes [10, 13]. Structural characteristic is a feature that people may identify, whether number of bedrooms and bathrooms, or floor space, or garage and patio. These structural attributes, often offered by house builders or developers to attract potential buyers, therefore meet the potential buyers' wishes. In [14] in his earlier study, structural attributes would be the key consideration for house hunters in determining what to purchase as such attributes represent their market value. In their earlier study, [15] stated that all these attributes have a positive relationship to rising house prices [16].

C. Neighbourhood

Neighborhood qualities can be included in deciding house price. According to [13], efficiency of public education, community social status and proximity to shopping malls typically improve the worth of a property. There is a substantial rise in house prices from the fifth-class suburban community to affluent neighborhood as predicted [16]. Nonetheless, [13] study found that these qualities tend to be cultural based, as they are not similarly relevant in all cultures.

Table 1. Summarization of Attributes used in Previous Study

Previous Study		[6]	[3]	[4]	[17]	[18]	[19]	[5]	[10]	[9]	Data Source
Locational Attributes	Access to shopping mall	✓	✓	✓	✓	✓		✓	✓		7
	Access to schools	✓			✓			✓	✓	✓	5
	Access to hospitals	✓	✓	✓	✓			✓	✓	✓	7
	Restaurants				✓			✓	✓	✓	4
	Public Transportation	✓	✓		✓			✓	✓	✓	6
Structural Attributes	No of bedroom	✓	✓	✓	✓			✓			5
	No of bathroom		✓			✓	✓				3
	Floor area	✓			✓	✓					3
	Garage and patio				✓	✓	✓				3
	Property age housing		✓	✓	✓	✓					4
	Lot size		✓	✓	✓	✓	✓	✓	✓		7
Neighborhood Attributes	Socio-economic variables				✓	✓				✓	3
	Local government										0
	Crime rates										0
	Place of worship							✓	✓		2
	Pleasant landscape		✓					✓			2
	Quiet atmosphere		✓								1
Economic Attributes	Income									✓	1
	Cost of material									✓	1

Based on Table 1, it is cleared that locational attributes which consist of access to shopping mall, access to school, access to hospital, restaurants and public transport are the most common attributes used in prediction of house price followed by neighborhood and economic attributes.

3. Machine Learning Model

According to [20], the paradigm of evaluating the house demand can be classified into two classes which are the traditional method and the advanced valuation method. The traditional valuation scheme, including multiple regression method and stepwise regression process, whilst hedonic pricing tool, artificial neural network (ANN) and spatial analysis framework are advances valuation method. The model selection to be used to predict house price is quite critical as varieties of models are available. One of the most commonly utilized models in this research field is Regression Analysis which is used in many studies, including [3, 10, 21]. Another common model for house price predictions is the Support Vector Regression (SVR) [7, 22, 23].

A. Regression Analysis

i. Hedonic Price Model

The housing market is slightly different from normal good consumption. According to [13], housing market is unique because it displays the characteristics of resilience, flexibility and spatial fixity. Therefore, hedonic approach is preferred to accurately predict market differential. [24] conceived hedonic model back in 1939, but this research was popularized in the early 1960s with comprehensive uses by Zvi Griliches and Rosen [24]. In the early 1930s, Court used this model to analyse automotive value in pricing and quality characteristics. [25] defined hedonic as "the implicit prices of attributes and are revealed to economic agents from the observed prices of differentiated products and the specific quantities of their characteristics." Following years of progress, Rosen applied the approach to the residential home price study and became commonly included in real estate sector research [3].

Rosen's philosophy or model comprises two separate phases. The regression of a product price on its attributes is performed in the initial stage to calculate the aggregate price of the component. A measure of a goods price will be determined in the first stage, but the inverse demand function cannot be generated at this stage. Thus, the second stage of estimation is needed to identify the inverse demand function that can be derived from the first stage implicit price function. In an earlier study, a study compared three commonly used house price measurement methods which are

simple average method, hedonic model, and matching approach. The result found that when adopted on the housing market, two methods that are simple average method and matching approach were proven biased. Thus, the hedonic model provides the highest results relative to those two most commonly encountered versions [3].

Hedonic pricing model is a statistical model that believes the worth of the property is the sum of all its attributes based on hedonic market theory.

ii. Multiple Linear Regression

Regression analysis is a model used to determine the relationship between variables. In order to evaluate the correlation of the variables, the correlation coefficient or regression equation can be used [26]. Multiple regression models can determine which characteristics are the most important to explain the dependent variable. Multiple regression analysis also allows certain price predictions by capturing independent and dependent variable data. In [27], the power of the multiple regression model can be seen when the value of the relationship between dependent and independent variables is measured. [28] use multiple regression modelling to describe improvements to an independent variable with a dependent variable.

This model can be achieved using the house price projection as separate and dependent variables like house prices, house size, property sort, number of bedrooms, and many more. Therefore, the house price is set as a target or dependency variable, while other attributes are set as independent variables to determine the main variables by identifying the correlation coefficient of each attribute.

B. Support Vector Regression

Support vector regression is a predictive model based on SVM, a neural network that usually has three layers, a powerful form of supervised learning. The model is based on a subset of training data. The advantages of support vector regression are that it is capable of processing non-linear results, provides only one possible optimal solution, and able to overcome a small sample learning issues [23].

The potential to produce market predictions in several markets, including real estate, shows that this model can overcome the non-linear regression problems and small sample learning problems. Moreover, as this model did not depend on probability distribution assumptions, and the ability of mapping the input attribute, either linear or non-linear, this model was commonly used at house price modelling [22]. Support vector regression offers huge benefits in so many aspects as this model can avoid over-fitting problems, while ensuring a single optimum solution by minimizing structural risks and empirical risks [29].

In this field of study, support vector regression is used to collect details on neighborhood, structural and locational attributes.

C. Artificial neural Network

In 1958, [30] created artificial neural network known as ANN. Walter Pitts and Warren McCulloch published a paper entitled "A Logical Calculus of Ideas Immanent in Nervous Activity" in the year 1943 which notes that a neural network may artificially be created, based on the role and structure of a biological neural network. In another research, as this model would often promote learning, artificial neural networks are claimed to be artificial brain diagrams [31, 32].

The artificial neural network model has always been selected when a non-linear attribute is involved. The analysis of home price estimation should also use this model as a spatial consideration for the price of housing is also non-linear. Therefore, as in [32-35], their study produces a good result, thus it is promising to provide an exact predictive model utilizing the artificial neural network algorithm. This system, however, has very limited performance. ANN can model complex non-linear relationships as house price predictions involve many non-linear variables.

D. Gradient Boost

Gradient boosting was created by [36] in 1999 and is a commonly used machine learning algorithm because of its performance, consistency and interpretability. Gradient boosting delivers state-of-the-art in various machine learning activities, such as multistage classification, click prediction and ranking. With the advent of big data in recent years, gradient boosting faces new challenges, especially with regard to the balance between accuracy and performance [37]. There are few parameters for gradient boosting. To ensure a dynamic balance between fit and regularity, the following steps can be taken to select parameters: (1) Setting regularization parameters (λ , α), (2) reducing learning rate and decide those optimal parameters again [19].

4. Related Work

A total of 14 articles were reviewed and evaluated to capture all attributes that influences the price of house. [3] in his article stated that square footage of unit of a house is the most importance variable in predicting price of a house, followed by number of bathrooms and number of bedrooms. Apart from that, the study suggests that the worth of the house increases by 2.6% if the floor space of the house is raised by 100 square feet. They also conclude that when the

building's operating year is 1 year lower, the value increase by 0.3 percent. In addition, the price of a house would increase by 10.4 or 13.7 percent, with one more bedroom or one more bathroom.

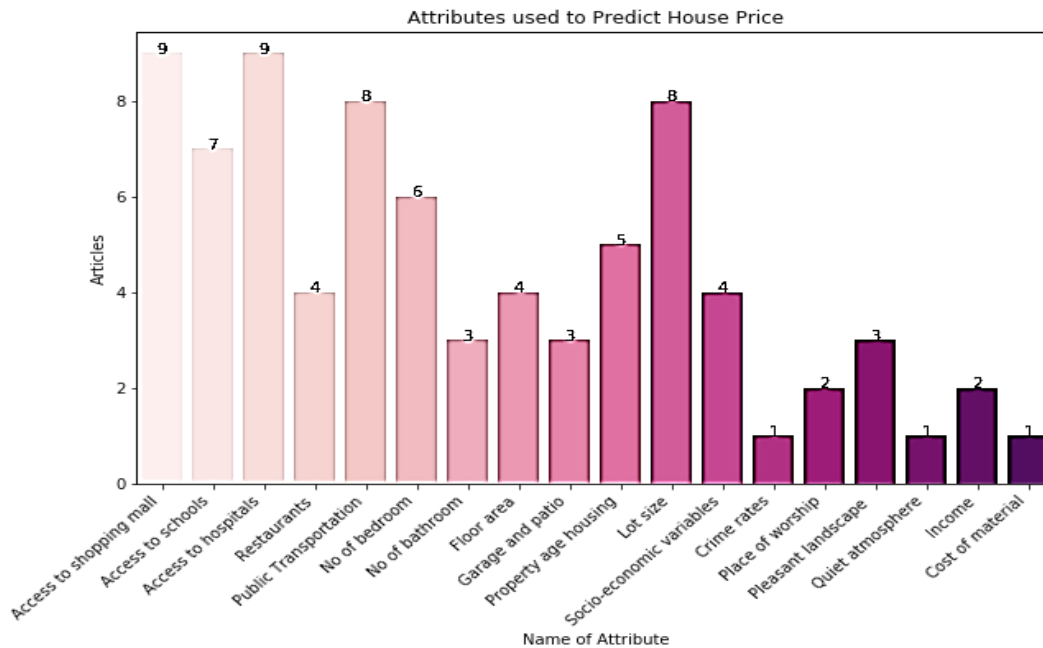


Fig.1. Types of Attributes used in Previous Study

From previous study reviewed, 19 attributes are stated to have been widely used by other researchers to evaluate house price. The attributes from 12 articles, which is shown in Fig. 1, are both grouped and tabled in a bar graph. It is therefore obvious what attributes researchers have used extensively to determine house prices. The number in the above bar graph represents the cumulative papers that used the attributes as predictor.

The diagram above shows that a shopping mall, hospital access and a house's lot size are the main attributes used to establish house prices. The research in recent years has dominated debates concerning locational attributes, including access to shopping centers and hospitals, as well as structural attributes, including number of bedrooms and the lot size of the house.

In fact, [16] described the location as an important house price predictor. The contribution of location attribute to house price drops as predicted from first class residential districts to fifth class residential districts. [5] pointed out that the four objects which most affected the house price are hospitals, schools, campuses, and leisure parks, which can be included in the locational attributes.

In comparison, 8 of 14 studies have used the structural attributes to determine house prices, including number of bedrooms, number of bathrooms and the lot size of the house. [38] clarified that the three major attributes that influence home selling rates are the total square feet, overall efficiency and the total number of bathroom units. This finding is in line with [7] who discovered that the number of bedrooms and bathrooms and the home price have significant relations. In a similar vein, [17] stressed that an added floor, bedroom and washroom add 13 per cent, 16 per cent and 2 per cent to the home price, respectively.

In addition to the locational and structural attributes, many researchers use the attributes of neighborhood to determine house prices. This can be seen in [12], where he claimed that neighborhood influences affected the house price, because citizens are likely to select a better neighborhood today. The neighborhood attributes also include low crime rates, pleasant scenery and quiet atmosphere. These factors will determine whether the price of house is high or low.

Although only a few researchers chose economic attribute, including individual income and the expenses of constructing a house, as the factor in determining a house price, we agree that economic attributes do have a major impact on house prices. [9] stated in his study that a house price can be determined based on an individual income because the government plays a role in setting a house price dependent on individual financial conditions. [12] supports this study in saying that the relationship between house price and income is important to describe the affordability of a house. This is one of the factors leading to each person's affordability to own or rent a house. An evaluation of the main attributes that impact the house price is crucial and related to the first research question of this study.

After evaluating the main attributes impacting house price determination, the data mining methodology (which is in the context of this study is a predicting model) can be used to estimate house price.

In order to forecast house price, predictive modeling such as support vector regression or artificial neural network were used. Predictive modeling uses data mining to forecast what it observed during the study phase. Fig. 2 shows types

of predictive model that is used by researchers in previous study. However, four popular prediction models, or more known as classifiers, used by researchers to construct this predictive model which are multiple linear regression, supporting vector regression, artificial neural network and classifier gradient booster.

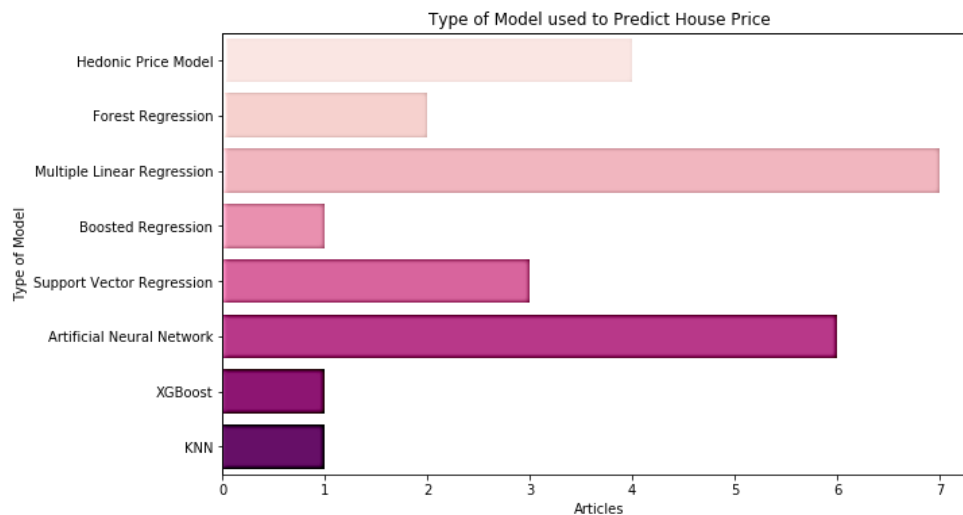


Fig 2. Types of Predictive Model used in Previous Study

Table 2. Summarization of Predictive Model used in Previous Study

Previous study	[38]	[3]	[7]	[39]	[34]	[32]	[40]	[22]	[10]	[19]	[41]	[11]	[23]	Data Source
Multiple Regression Analysis	✓	✓		✓			✓		✓		✓	✓		7
Support Vector Regression			✓					✓					✓	3
Artificial Neural Network			✓		✓	✓	✓						✓	5
XGBoost	✓									✓				2

Table 2 shows that the most popular model used to predict house price is by using multiple regression analysis. The hedonic price model was typically used for classification of important variables together with other regression model, such as support vector regression, multiple regression analysis and other models. Meanwhile, research by [38] selects XGBoost as the best model since it provides the lowest RMSE value in contrast with other models in his study. Such an analysis is related to the second research question of this study.

5. Finding and Discussion

The associations between the house price and predicting model included in this segment have been explored. In addition, the impact of various attributes on specific model have also been evaluated and debated.

Table 3. Analysis of House Attributes and RMSE of Each Model

Models	Attributes	RMSE	Previous Study
Multiple Linear Regression	Structural attributes	0.1261	[38]
	Locational and structural attributes	-	[3]
	Locational and structural attributes	0.267	[39]
	Locational attributes	-	[40]
	Locational attributes	-	[10]
	Economic attributes	-	[41]
	Locational, structural and neighborhood attribute	-	[11]
	Locational and structural attributes	0.3079	[7]

Support Vector Regression	Locational and structural attributes	0.2362	[7]
	Locational and structural attributes	-	[22]
	Locational attributes	0.0047	[23]
Artificial Neural Network	Locational and structural attributes	0.5155	[7]
	Locational attributes	-	[32]
	Locational attribute	-	[40]
	Locational attributes	0.0581	[23]
XGBoost	Structural attribute	0.1212	[38]

Based on reviewing numerous papers, there are several attributes used by researchers in their work to forecast house prices. All of these attributes can be divided into 4 main categories which are locational, structural, neighborhood and economic attributes. The locational attribute consists of variables which described the accessibility to shopping mall, accessibility to school, accessibility to hospital, restaurants and availability of public transport. Meanwhile, the structural attribute consists variables of number of bedrooms, number of bathrooms, floor area, garage and patio, property age housing and lot size. The neighborhood on the other hand described the socioeconomic variables, crime rates, place of worship, pleasant landscape, and quiet atmosphere which the variables is quite subjective. And lastly the economic attributes consist of income factor and cost of material factor. The most common used attributes by previous research regarding house price prediction are locational and structural attributes, as neighborhood and economic attributes are difficult to define and measure.

The classification of attributes made it easier to analyze the effects of different attributes on different models. The tabulation of finding based on evaluation of previous study can be seen in Table 3. Looking at the table above, it can be seen that support vector regression has the smallest RMSE value, 0.0047. Technically, the RMSE value of a model is highly dependent on the attributes used during the prediction process. Most of the model that are using the same attributes (locational attributes) will generate a very low RMSE value indicating the best model. However, this cannot simply show the best model because several previous studies rarely provide the RMSE value to justify their model being the best model. On the basis of the analysis table, the locational attribute may be assumed to be the main attribute used by several models such as support vector regression and artificial neural network support. The RMSE value is very low with the presence of the locational attribute only, however, the RMSE value is quite high when the structural attribute is combined with the locational attribute for the input to make a prediction.

Next, ANN provides the second lowest of RMSE value which is 0.0581. Findings revealed that locational attributes are indeed the relevant attributes used in the ANN model to forecast house price. Similar to SVR, the RMSE value is fairly high when locational and structural attributes are used together. This indicates that the locational attribute should be used alone to achieve low RMSE values by ANN model. In the meantime, the XGBoost model also yields the lowest RMSE while only the structural attribute is involved.

However, research on model with structural attributes alone is very limited as previous research focused mainly on locational attributes or the combination of locational and structural attributes in order to predict house prices. In general, our analysis suggests that SVR, ANN and XGBoost are the most efficient models compared to other models, whereas locational attributes are the main attribute in predicting house price.

6. Conclusions

This paper examined and analyzed the current research on the significant attributes of house price and analyzed the data mining techniques used to predict house price. Technically, houses with a strategic location such as the accessibility to shopping mall or other facilities tend to be more expensive than houses in rural areas with limited numbers of facilities.

The accurate prediction model would allow investors or house buyers to determine the realistic price of a house as well as the house developers to decide the affordable house price. This paper addressed the attributes used by previous researchers to forecast a house price using various prediction models. Taken together, the results of the survey have shown the potential of SVR, ANN and XGBoost in predicting house prices. These models were developed based on several input attributes and they work significantly positive with house price. In conclusion, the impact of this research was intended to help and assist other researchers in developing a real model which can easily and accurately predict house prices. Further work on a real model needs to be done with the utilization of our findings to confirm them.

References

- [1] A. S. Temür, M. Akgün, and G. Temür, "Predicting Housing Sales in Turkey Using Arima, Lstm and Hybrid Models," *J. Bus. Econ. Manag.*, vol. 20, no. 5, pp. 920–938, 2019, doi: 10.3846/jbem.2019.10190.
- [2] A. Ebekozien, A. R. Abdul-Aziz, and M. Jaafar, "Housing finance inaccessibility for low-income earners in Malaysia: Factors and solutions," *Habitat Int.*, vol. 87, no. April, pp. 27–35, 2019, doi: 10.1016/j.habitatint.2019.03.009.
- [3] A. Jafari and R. Akhavian, "Driving forces for the US residential housing price: a predictive analysis," *Built Environ. Proj. Asset Manag.*, vol. 9, no. 4, pp. 515–529, 2019, doi: 10.1108/BEPAM-07-2018-0100.
- [4] Choong Wei Cheng, "Statistical Analysis of Housing Prices in Petaling," Universiti Tunku Abdul Rahman, 2018.

- [5] R. E. Febrita, A. N. Alfiyatin, H. Taufiq, and W. F. Mahmudy, "Data-driven fuzzy rule extraction for housing price prediction in Malang, East Java," *2017 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSIS 2017*, vol. 2018-Janua, pp. 351–358, 2018, doi: 10.1109/ICACSIS.2017.8355058.
- [6] G. Gao *et al.*, "Location-Centered House Price Prediction: A Multi-Task Learning Approach," pp. 1–14, 2019, [Online]. Available: <http://arxiv.org/abs/1901.01774>.
- [7] T. D. Phan, "Housing price prediction using machine learning algorithms: The case of Melbourne city, Australia," *Proc. - Int. Conf. Mach. Learn. Data Eng. iCMLDE 2018*, pp. 8–13, 2019, doi: 10.1109/iCMLDE.2018.00017.
- [8] Y. Y. S. Song, T. Zhou, H. Yachi, and S. Gao, "Forecasting house price index of China using dendritic neuron model," *PIC 2016 - Proc. 2016 IEEE Int. Conf. Prog. Informatics Comput.*, pp. 37–41, 2017, doi: 10.1109/PIC.2016.7949463.
- [9] R. Aswin Rahadi, S. K. Wiryono, D. P. Koesrindartoto, and I. B. Syamwil, "Factors Affecting Housing Products Price in Jakarta Metropolitan Region," *Int. J. Prop. Sci.*, vol. 6, no. 1, pp. 1–21, 2016, doi: 10.22452/ijps.vol6no1.2.
- [10] A. Nur, R. Ema, H. Taufiq, and W. Firdaus, "Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization Case Study : Malang, East Java, Indonesia," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 10, pp. 323–326, 2017, doi: 10.14569/ijacs.2017.081042.
- [11] A. Yusof and S. Ismail, "Multiple Regressions in Analysing House Price Variations," *Commun. IBIMA*, vol. 2012, pp. 1–9, 2012, doi: 10.5171/2012.383101.
- [12] A. Osmadi, E. M. Kamal, H. Hassan, and H. A. Fattah, "Exploring the elements of housing price in Malaysia," *Asian Soc. Sci.*, vol. 11, no. 24, pp. 26–38, 2015, doi: 10.5539/ass.v11n24p26.
- [13] T. L. Chin and K. W. Chau, "A critical review of literature on the hedonic price model," *Int. J. Hous. Sci. Its Appl.*, vol. 27, no. 2, pp. 145–165, 2003.
- [14] M. J. Ball, "Recent Empirical Work on the Determinants of Relative House Prices," *Urban Stud.*, vol. 10, no. 2, pp. 213–233, 1973, doi: 10.1080/00420987320080311.
- [15] M. Rodriguez, "Managing Corporate Real Estate: Evidence from the Capital Markets." *Journal of Real Estate Literature*, 1996.
- [16] D. G. Owusu-Manu, D. J. Edwards, K. A. Donkor-Hyiaman, R. O. Asiedu, M. R. Hosseini, and E. Obiri-Yeboah, "Housing attributes and relative house prices in Ghana," *Int. J. Build. Pathol. Adapt.*, vol. 37, no. 5, pp. 733–746, 2019, doi: 10.1108/IJBPA-01-2019-0003.
- [17] D.-G. Owusu-Manu, D. J. Edwards, K. A. Donkor-Hyiaman, R. O. Asiedu, M. R. Hosseini, and E. Obiri-Yeboah, "Housing attributes and relative house prices in Ghana," *Int. J. Hous. Mark. Anal.*, vol. 8, no. 2, p. 1998, 2018, doi: 10.1017/CBO9781107415324.004.
- [18] J. M. Montero, R. Míguez, and G. Fernández-Avilés, "Housing price prediction: parametric versus semi-parametric spatial hedonic models," *J. Geogr. Syst.*, vol. 20, no. 1, pp. 27–55, 2018, doi: 10.1007/s10109-017-0257-y.
- [19] S. Lu, Z. Li, Z. Qin, X. Yang, and R. S. M. Goh, "A hybrid regression technique for house prices prediction," *IEEE Int. Conf. Ind. Eng. Eng. Manag.*, vol. 2017-Decem, pp. 319–323, 2018, doi: 10.1109/IEEM.2017.8289904.
- [20] 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, doi: 10.1108/14635780310483656.
- [21] Y. F. Chang, W. C. Choong, S. Y. Looi, W. Y. Pan, and H. L. Goh, "Analysis of housing prices in Petaling district, Malaysia using functional relationship model," *Int. J. Hous. Mark. Anal.*, vol. 12, no. 5, pp. 884–905, 2019, doi: 10.1108/ijhma-12-2018-0099.
- [22] J. H. Chen, C. F. Ong, L. Zheng, and S. C. Hsu, "Forecasting spatial dynamics of the housing market using Support Vector Machine," *Int. J. Strateg. Prop. Manag.*, vol. 21, no. 3, pp. 273–283, 2017, doi: 10.3846/1648715X.2016.1259190.
- [23] H. Y. Lin and K. Chen, "Predicting price of Taiwan real estates by neural networks and Support Vector Regression," *Recent Res. Syst. Sci. - Proc. 15th WSEAS Int. Conf. Syst. Part 15th WSEAS CSCC Multiconference*, pp. 220–225, 2011.
- [24] A. C. Goodman, "Andrew Court and the Invention of Hedonic Price Analysis," *J. Urban Econ.*, vol. 44, no. 2, pp. 291–298, 1998, doi: 10.1006/juec.1997.2071.
- [25] S. Rosen, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *J. of Political Econ.*, vol. 82, no. 50, pp. 34–55, 1974, doi: 10.1016/S0040-4039(00)85403-9.
- [26] J. Maxey, "The effect of pricing factors on real estate transactions in Prince George 's county , Maryland," no. February, p. 160, 2013, [Online]. Available: <http://search.proquest.com.ezproxy.apollolibrary.com/docview/1372292022?accountid=35812>.
- [27] S. P. Ellis and S. Morgenthaler, "Leverage and breakdown in L1regression," *J. Am. Stat. Assoc.*, vol. 87, no. 417, pp. 143–148, 1992, doi: 10.1080/01621459.1992.10475185.
- [28] P. Cohen, J. Cohen, J. Teresi, M. Marchi, and C. N. Velez, "Problems in the Measurement of Latent Variables in Structural Equations Causal Models," *Appl. Psychol. Meas.*, vol. 14, no. 2, pp. 183–196, 1990, doi: 10.1177/014662169001400207.
- [29] T.-W. Lee and K. Chen, "Prediction of House Unit Price in Taipei City Using Support Vector Regression," 2008, [Online]. Available: <http://apiems2016.conf.tw/site/userdata/1087/papers/0307.pdf>.
- [30] F. Rosenblatt, "Recent Work on Theoretical Models of Biological Memory." 1958.
- [31] P. Jaiswal, N. K. Gupta, and A. Ambikapathy, "Comparative study of various training algorithms of artificial neural network," *2018 Int. Conf. Adv. Comput. Commun. Control Netw.*, pp. 1097–1101, 2019, doi: 10.1109/icaccn.2018.8748660.
- [32] M. F. Mukhlshin, R. Saputra, and A. Wibowo, "Predicting house sale price using fuzzy logic, Artificial Neural Network and K-Nearest Neighbor," *Proc. - 2017 1st Int. Conf. Informatics Comput. Sci. ICICoS 2017*, vol. 2018-Janua, no. 1, pp. 171–176, 2018, doi: 10.1109/ICICOS.2017.8276357.
- [33] W. T. Lim, L. Wang, Y. Wang, and Q. Chang, "Housing price prediction using neural networks," *2016 12th Int. Conf. Nat. Comput. Fuzzy Syst. Knowl. Discov. ICNC-FSKD 2016*, pp. 518–522, 2016, doi: 10.1109/FSKD.2016.7603227.
- [34] J. J. Wang *et al.*, "Predicting House Price with a Memristor-Based Artificial Neural Network," *IEEE Access*, vol. 6, pp. 16523–16528, 2018, doi: 10.1109/ACCESS.2018.2814065.
- [35] H. Wu *et al.*, "Influence factors and regression model of urban housing prices based on internet open access data," *Sustain.*, vol. 10, no. 5, pp. 1–17, 2018, doi: 10.3390/su10051676.
- [36] J. H. Friedman, "Stochastic Gradient Boosting," vol. 1, no. 3, pp. 1–10, 1999.

- [37] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 3147–3155, 2017.
- [38] Y. Zhou, "Housing Sale Price Prediction Using Machine Learning Algorithms," 2020.
- [39] T. Mohd, S. Masrom, and N. Johari, "Machine learning housing price prediction in petaling jaya, Selangor, Malaysia," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2 Special Issue 11, pp. 542–546, 2019, doi: 10.35940/ijrte.B1084.0982S1119.
- [40] A. Varma, A. Sarma, S. Doshi, and R. Nair, "House Price Prediction Using Machine Learning and Neural Networks," *Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2018*, pp. 1936–1939, 2018, doi: 10.1109/ICICCT.2018.8473231.
- [41] R. Reed, "The relationship between house prices and demographic variables: An Australian case study," *Int. J. Hous. Mark. Anal.*, vol. 9, no. 4, pp. 520–537, 2016, doi: 10.1108/IJHMA-02-2016-0013.

Authors' Profiles



Nor Hamizah Zulkifley received her BSc. degree of Actuarial Science in 2019 from University Teknologi Mara. Currently, she is a master's degree student of Data Science at the Faculty of Computer and Mathematical Science in the same university.



Dr. Shuzlina Abdul Rahman is an Associate Professor of Information System in Universiti Teknologi Mara. She received a master's degree in information technology in Universiti Utara Malaysia in 2000. In 2012, she received her PhD degree from the Universiti Kebangsaan Malaysia. She teaches courses related to the philosophy of artificial intelligence, data mining, advanced decision support systems, business intelligence and intelligent systems development. Her primary research interests involve the computational intelligence, data mining and optimization and intelligent data analytics.



Dr. Nor Hasbiah Ubaidullah is an Associate Professor at Universiti Pendidikan Sultan Idris. She received degree in Computer Science from Universiti Kebangsaan Malaysia and Master of Science in Information System from University of Salford. She obtained her PhD degree in Information Technology from Universiti Kebangsaan Malaysia. She has been teaching in higher education institutions since 1991 in several subjects such as Programming, Courseware Engineering and Information Systems Development. Her research areas include Data Management, Educational Software and Computer Science Education.



Dr. Ismail Ibrahim is working as a lecturer at Universiti Teknologi Mara, Malaysia. He received a degree in Electrical Engineering from Universiti Sains Malaysia and a master's in electrical engineering. He obtained his PhD degree in Electrical Engineering from Universiti Malaysia Pahang. He has been working in the industry since 2005. His research areas include Computational Intelligence, Optimization, Image Processing and Artificial Intelligence.

How to cite this paper: Nor Hamizah Zulkifley, Shuzlina Abdul Rahman, Nor Hasbiah Ubaidullah, Ismail Ibrahim, " House Price Prediction using a Machine Learning Model: A Survey of Literature", *International Journal of Modern Education and Computer Science(IJMECS)*, Vol.12, No.6, pp. 46-54, 2020.DOI: 10.5815/ijmecs.2020.06.04