https://doi.org/10.37134/jcit.vol13.2.5.2023

Forecasting on House Price Index using Artificial Neural Network

Gangaieisvari Gobalkrishnan*, & Zahayu Md Yusof

School of Quantitative Sciences, Universiti Utara Malaysia 06010 Sintok, Kedah.
**Email: ganges2894@yahoo.com*

Abstract

Forecasting the residential property sector is a crucial component in the decision-making process for investors and government in supporting asset allocation, developing property finance plans and implementing a relevant policy. The purpose of this study is to examine the determinants of Penang house price index and to develop a model to forecast Penang house price index in Malaysia. Estimation is done by using ordinary least square and artificial neural network method. Relevant data sets were obtained from the Monthly Statistical Bulletin, Bank Negara Malaysia and National Property Information Centre. The empirical analysis of this research is based on quarterly time series data which cover the periods from 2005Q1 to 2022Q1. The main findings reported that base lending rate and unemployment rate are negatively associated with and have significant impacts on Penang house price index. Meanwhile, gross domestic product is positively related to and has a significant impact on Penang house price index. Consumer price index shows a positive sign; however, it recorded an insignificant impact on Penang house price index. Even though there are three independent variables recorded significant impact on Penang house price index, yet gross domestic product is the most vital determinant of Penang house price index in Malaysia. The artificial neural network model was trained and tested using quarterly time series data from 2005Q1 to 2022Q1 and the model was validated using data from 2021Q1 to 2022Q1. Model validation indicates that artificial neural network has a high level of accuracy in its ability to learn, generalize, and converge time series data efficiently as well as able to generate reliable forecasting information.

Keywords: Artificial neural network; House price index

1. Introduction

One of the most fundamental needs of all humans is the need for a house (Forbes, 2018). When it comes to shelter or accommodation, it is defined as a place that provides living space for individuals or families as well as provides protection from extreme weather or danger. Based on NAPIC (2021), the average property price in Penang is reportedly RM 437,632. Penang property prices are slightly higher than the national average, which is around RM 420,345, but still far below the Kuala Lumpur average of RM 780,564 (Gregor, 2020).

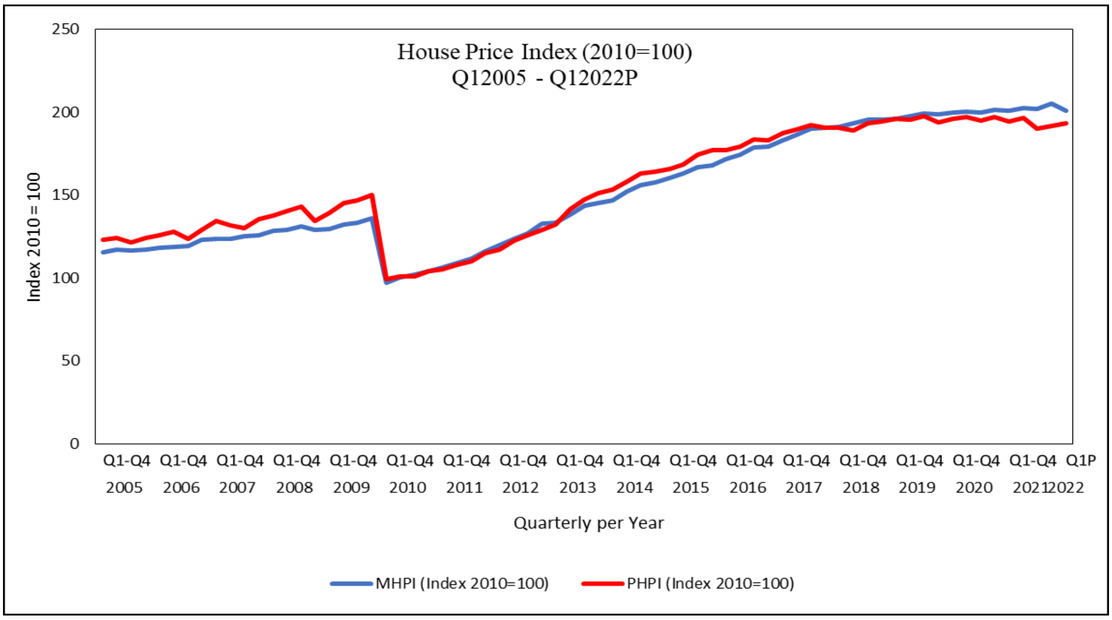

Figure 1: Malaysia house price index (MHPI) and Penang house price index (PHPI) from 2005Q1 to 2022Q1

Over the above, the pattern of MHPI from year 2005Q1 to 2009Q4 shows a continuous rise from index of 115.4 to index of 136.1 and eventually shows a decreasing trend at year 2010 (Q1-Q4) due to the base year has been structured from year 2000 to 2010 where the index number is equals to 100.0 in base year 2010 (NAPIC, 2022). Thereafter, the MHPI continuous to shows an increasing trend from year 2011Q1 with index of 106.4 till 2022Q1 with index of 200.9. Similarly, the pattern for PHPI from 2005Q1 to 2009Q4 shows a continuous rise from index of 122.9 to index of 149.9 and thereafter shows a decreasing trend at year 2010 (Q1-Q4) due to base year amendments and then continues to shows an increasing trend from year 2011Q1 with index of 105.1 till 2022Q1P with index of 193.5. Many people are concerned about the challenges of owning a house because of this growth especially those in the B40 and M40 income levels. In addition, the construction industry in Malaysia has been navigating through COVID-19 and all the consequences have risen from it. In 2020, construction start-ups to fell by 19.39 % to 58.28 %, owning to delays in several major development projects and surge in unsold housing stock (Construction Plus Asia, 2022).

Moreover, understanding of Penang state house price index (HPI) trends is important for investor, economists, practitioners, and even private individuals (Radzi et al., 2012). Therefore, forecasting HPI would enable home buyers and the developers to make a crucial and suitable decision to reduce the risk of loss contributing to the house prices and the unsold properties. In this study, there are three objectives need to achieve which are to identify the determinants of PHPI, to examine the most vital determinant of PHPI and to develop a model to forecast PHPI in Malaysia.

2. Literature Review

Concept of HPI

MHPI is created by Valuation and Property Service Department (VPSD) and published by NAPIC on a quarterly and annual basis (Rahman and Ridzuan, 2020). Sukrri et al. (2019) indicates that the MHPI has a substantial long-term association with employment, overnight policy rate, consumer price index, land availability and housing loans but not with building costs. Meanwhile, Garg (2016) investigated HPI by using econometric methods, artificial neural network (ANN) models, and time series model such as auto regression. The findings reveal that interest rate, oil price shocks, currency movement, business cycle, legislative and regulatory policies, and inflation are some of the important macroeconomic factors used in modelling and forecasting HPI.

Response of HPI with the macroeconomic variables such as gross domestic product (GDP), base lending rate (BLR), consumer price index (CPI) and unemployment rate (UR) will be present in Table 1.

Table 1: Determinants of HPI

Variables	Descriptions	Author(s)
	• Conducted study on the long-term relationship and causality effect of the HPI and its factors. Findings revealed that there is a long-run relationship and unidirectional link between HPI and GDP.	Jehani et al. (2020)
BLR	• Conducted a study on the impact of macroeconomic indicators on HPI for the Town of Amherst, New York, USA by using Vector autoregression approach. The findings reveal that mortgage interest rate and HPI are positively related and statistically significant.	Mohan et al. (2019)
	• Determined the most influential elements influencing HPI in Malaysia by using autoregressive distributed lag (ARDL). The results show that BLR is significant and has a positive relationship with the MHPI, which was also supported by Tumbarello & Wang (2010) who noted that BLR is statistically significant and positively correlated with MHPI. This is because when BLR rises, housing developers are required to borrow at a higher cost.	Rahman and Ridzuan (2020); Tumbarello and Wang (2010)
CPI	• Investigated the determinants of HPI in Malaysia and Singapore by using multiple linear regression analysis. The findings indicated that there is a positive correlation between the inflation rate and HPI in Malaysia and a negative correlation with HPI in Singapore.	Lee and Azlan (2022); Trofimov et al. (2018)
UR	• Examined the influence of house price on UR and stock market in China from 2010 to 2019. The result reveals that house price on unemployment is positively correlated in the long run whereas in the short term, it is negatively correlated.	Sun (2021)
	• This researcher used house price data from 1970 to 2010 for 34 countries on a yearly basis. The findings revealed that there is an inverse relationship between the house price and UR by using ordinary least square (OLS) and two-stage least square approach.	Geerolf and Grjebine (2014)

Application of ANN Model in Forecasting HPI

Ge et al. (2021) examined a Ghanaian case study on the influence of artificial intelligence real estate forecasting using multiple regression and ANN using data from Ghanaian apartment auctions from 2016 to 2020. The ANN model performs the best performance and effective zonal segmentation based on auction evaluation price significantly improve the model's prediction accuracy. Meanwhile, Kitapci et al. (2017) conducted research to forecast housing price in Ankara, Turkey from January 2010 to September 2016 using an ANN technique. The findings showed that the proposed model has 78% of success rate and that approach may be used to help in decision-making particularly for real estate investors and final consumers to forecast the future housing price in Ankara.

3. Methodology

Figure 2 shows the flow chart illustrating the research approach of this study. To achieve the research objective 1 and 2, this study will be conducted through OLS estimation method. Moreover, to achieve the research objective 3, ANN method will be conducted.

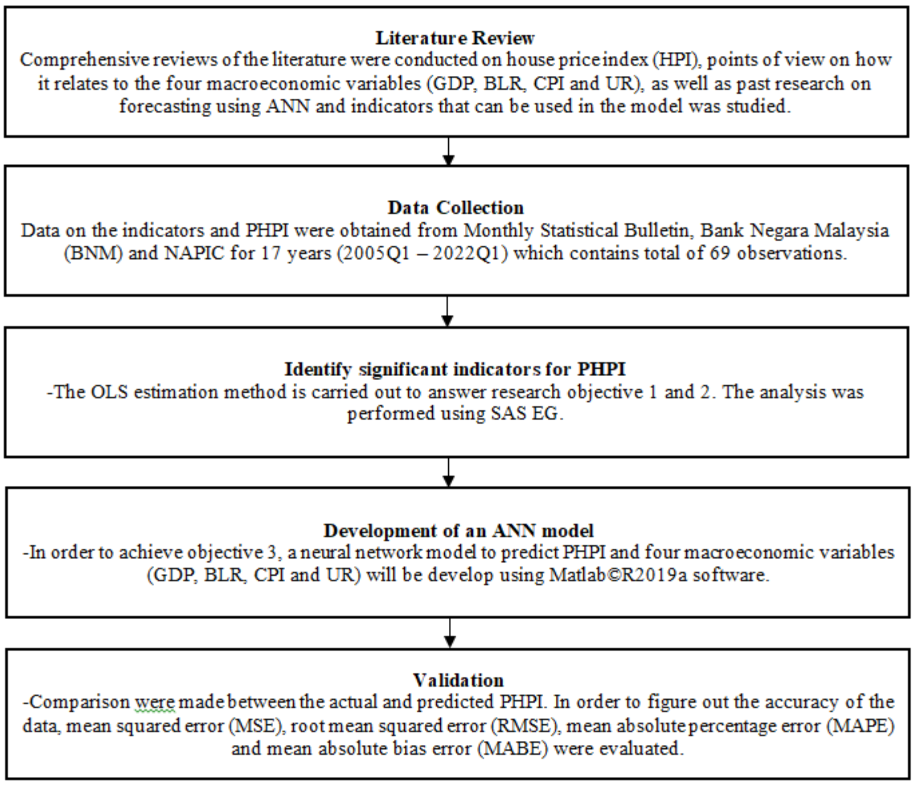

Figure 2: A flow chart illustrating the research approach

Research Framework

The research framework structure is derived from the selected literature and illustrated based on the research objectives. Figure 3 illustrates the determinants that influence PHPI in Malaysia.

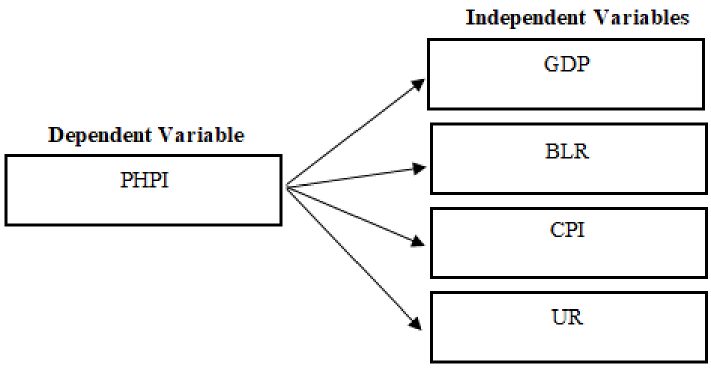

Figure 3: Determinants that influence PHPI

Methods of Data Analysis

Unit root test will be conducted prior to all other tests to ensure that the variables are in a stationary state. OLS estimation is used to analyse the relationship between PHPI in Malaysia and independent variables over the quarterly period from 2005Q1 to 2022Q1. By using SAS EG software, the OLS estimation method is carried out to answer research objective 1 and 2. A set of tests are conducted for diagnostic checking to detect the presence of multicollinearity and autocorrelation respectively. Meanwhile, research objective 3 where the application of ANN in developing the model for forecasting PHPI will be developed using Matlab version 2019a software and generated by training, testing, and validating process within the neurons of the ANN.

4. Result and Discussion

Descriptive Data Analysis and Findings

Table 2 displays the results of SAS EG computation of the number of observations, mean, median, standard deviation, skewness, and kurtosis values of the variables.

Table 2: Descriptive statistics

Variables	Mean	Median	Standard Deviation	Skewness	Kurtosis
*ln*PHPI	2.2462	2.2800	0.066	-1.4780	0.8006
*ln*GDP	5.3480	5.3800	0.1809	-0.4872	-1.2873
BLR	6.3986	6.5400	0.4630	-1.0200	-0.3821
*ln*CPI	2.0464	2.0500	0.0305	0.0103	-1.2736
UR	3.4380	3.4000	0.5102	1.7500	2.7598

Note: *ln* implies logarithm form.

Based on Table 2, it can be noticed that BLR has the highest mean and median which are 6.3986 and 6.5400 respectively. All the variables in this study have a standard deviation less than one, indicating that the data points tend to be close to the mean of the data set (Narkhede, 2018). Moreover, the descriptive statistics show that *ln*PHPI, *ln*GDP and BLR are having negative values for the skewness which indicates the data are skewed to the left at -1.4780, -0.4872 and -1.0200. In contrast, *ln*CPI and UR are having positive values for the skewness which indicates the data are skewed to the right at 0.0103 and 1.7500 respectively. Since the kurtosis values for all the variables in this study are less than 3 which are called as platykurtic indicates that the variables are not volatile.

Unit Root Test

To prove the stationarity status, this study employs Augmented Dickey-Fuller (ADF) unit root test for all the variables at level and first difference to determine the series order of integration which is stationary for the series.

Table 3: Augmented Dickey-Fuller unit root test result

Variables	*Level*		*First difference*	
	Intercept	Trend & Intercept	Intercept	Trend & Intercept
*ln*PHPI	-3.4690*** (0.0119)	-1.5386 (0.8062)	-8.27729*** (0.0000)	-9.3951*** (0.0000)
*ln*GDP	-1.3351 (0.6086)	-1.8900 (0.6488)	-8.6172*** (0.0000)	-8.6455*** (0.0000)
BLR	-2.7306* (0.0742)	-2.7295 (0.2286)	-4.1813*** (0.0014)	-5.0091*** (0.0006)
*ln*CPI	-1.2683 (0.6397)	-2.3826 (0.3851)	-8.6404*** (0.0000)	-8.5658*** (0.0000)
UR	-2.1918 (0.211)	-2.6836 (0.2466)	-9.0512*** (0.0000)	-9.0340*** (0.0000)

Notes: *ln* implies logarithm form. Figures in parentheses (…) are t-statistics. *** Significant at 1% significant level; ** Significant at 5% significant level; * Significant at 10% significant level.

Table 3 presents the results of ADF unit root test for five variables at level and first difference. The unit root test was tested in level form at first for both intercept and trend & intercept. The test was then repeated with the first difference form instead of the level form to differentiate the non-stationary and transform to stationary data. The optimal number of lags is based on the minimum SIC (Schwarz Information Criterion). ADF test shows that *ln*GDP, *ln*CPI and UR appear to be non-stationary for both intercept and trend & intercept at level. However, the result show that *ln*PHPI and BLR turn into stationary only under intercept at level where it's p-value at 0.0119 and 0.0742 are statistically significant at 1% and 10% significant level. Therefore, it can conclude that unit root is still existed in this model due to the non-stationary result of *ln*PHPI and BLR under trend & intercept at level whereas for *ln*GDP, *ln*CPI and UR are non-stationary under both intercept and trend & intercept at level. For that reason, ADF unit root test has been proceeded to the first difference form. The findings in Table 3 clearly shows that when ADF unit root test proceeds to first difference, all the variables can reject null hypothesis as their p-value is less than significance levels at 1%, 5% and even 10% significant levels. At this point, all the variables revealed that they were stationary at the first difference level, therefore it is possible to draw the conclusion that this model does not include any unit roots.

The Determinants of PHPI in Malaysia

Table 4 displayed the summary results of OLS estimation on how the selected independent variables (*ln*GDP, BLR, *ln*CPI and UR) impacts the *ln*PHPI in Malaysia.

Table 4: The determinants of PHPI in Malaysia

	Coefficient	*p*-value	Standard error
*ln*GDP	0.3361 (10.74)	0.0000***	0.0313
BLR	-0.0491 (-3.67)	0.0005***	0.0134
*ln*CPI	0.1241 (0.58)	0.5639	0.2138
UR	-0.0552 (-4.06)	0.0001***	0.01362
Constant	0.6994 (2.31)	0.0240**	0.3024
R-squared (R^2)	0.7858	-	-
Adjusted R-squared ($\bar{R}^2$)	0.7724	-	-
F-statistic	58.70	-	-

Notes: ln implies logarithm form. Figures in parentheses (…) are t-statistics. *** Significant at 1% significant level; ** Significant at 5% significant level; * Significant at 10% significant level. The model of OLS estimation results is expressed as follow:

$$ln\widehat{PHPI}_t = \hat{\beta}_0 + \hat{\beta}_1 lnGDP_t + \hat{\beta}_2 BLR_t + \hat{\beta}_3 lnCPI_t + \hat{\beta}_4 UR_t + \mu_t$$
$$ln\widehat{PHPI}_t = 0.6994 + 0.3361\, lnGDP_t - 0.0491\, BLR_t + 0.1241\, lnCPI_t - 0.0552 UR_t + \mu_t$$

Sample, N = 69 Observations
t = 2005Q1 – 2022Q1 (Quarterly)
μ = Error Team
t = Time trend
All variables using log form except BLR and UR, because it is in percentage form (%)

The results showed that value of intercept, β_0 is 0.6994. It means the autonomous *ln*PHPI in Malaysia is 0.70% when the other independent variables are set as zero. If there are no other independent variables, the intercept in this equation is meaningless. Moreover, β_1 coefficient value is 0.3361, which indicates 1% rise in *ln*GDP, *ln*PHPI in Malaysia will tend to increase by 0.34% with all other independent variables held fixed. *ln*PHPI and *ln*GDP are statistically significant and positively related. Higher GDP will eventually create more jobs opportunities for citizen and raise income level, promoting a better standard living of household (Latif et al., 2020). β_2 coefficient value is -0.0491, this implies 1% increase in BLR, *ln*PHPI in Malaysia tends to reduce by 0.05% with all other independent variables held fixed. *ln*PHPI and BLR are statistically significant and inversely related. When the mortgage rates raise, the affordability of purchasing a house fall, resulting in a decline in demand and a fall in HPI (Jehani et al., 2020). β_3 coefficient value is 0.1241, this implied 1% increment in lnCPI, lnPHPI in Malaysia will tends to increase by 0.12% with all other independent variables held fixed. lnPHPI and lnCPI are statistically insignificant and positively related. When an inflation occurs, the prices of raw material for building a house will increase and drives up the construction costs. Thus, a developer who seeks profit would increase the selling price of a house in order to cover the increment in construction costs. The positive relationship and insignificant impact are in agreement with the studies conducted by Tan (2011). Tan (2011) revealed that the inflation rates have moderate negative value with housing price or in other words, inflation rate is not significant determinant of HPI. β_4 coefficient value is -0.0552, this implied that when UR increase 1%, PHPI will then reduce by 0.06% with all other independent variables held fixed. PHPI and UR are statistically significant and negatively related. Sun (2021) claimed that negative relationship occur due to the house prices in Malaysia are considered to be seriously unaffordable as the median all house price is relatively higher than the annual medium income. In terms of goodness of fit, the computed $R^2$of the model is 0.7858. It indicates that 78.58% changes in PHPI can be explained by the selected independent variables, while the remaining 21.42% can be explained by variables other than GDP, BLR, ICPI and UR, or simply the error term.

GDP as the Most Vital Determinants of PHPI in Malaysia

Based on Table 4, the results show that lnGDP, BLR and UR are statistically significant at 1% significant level whereas lnCPI is statistically insignificant at 1%, 5% and 10% significance level, yet the most vital determinant of PHPI in Malaysia is lnGDP. This is because a 1% change in lnGDP will cause 0.34% change in PHPI. In other words, the degree of elasticity of PHPI to lnGDP is elastic as 1% increase in lnGDP will tends to lead to an increase of 0.34% in PHPI in Malaysia. The degree of impact of lnGDP is relatively larger as compared to the other variables and this result is in accordance with Latif et al. (2020), Jehani et al. (2020) findings.

Multicollinearity Test Outcome

Table 5: Results of variance inflation factor

Variable	VIF
Constant	NA
lnGDP	2.22
BLR	2.66
lnCPI	2.96
UR	3.35
Mean VIF	2.80

Based on the VIF results reported in Table 5, it can be concluded that the multicollinearity problem does not exist in this model. This is because the value of VIF for every single variable is less than 10, where the value of VIF for lnGDP, BLR, lnCPI and UR are 2.22, 2.66, 2.96 and 3.35 respectively. In short, the model does not suffer from multicollinearity problem.

Autocorrelation Test Outcome

Durbin-Watson d test was employed to detect the presence of autocorrelation. From the Durbin-Watson significance table, 5% significance point for n=69 and k=4 shows that the critical d_L and d_U values are 1.47 and 1.73 respectively. Since the computed Durbin-Watson d statistic is 0.463 which lies in the region $0 < d < d_L$ as shown in Figure 4, thus, do not reject H_0 as the model show the occurrence of positive autocorrelation.

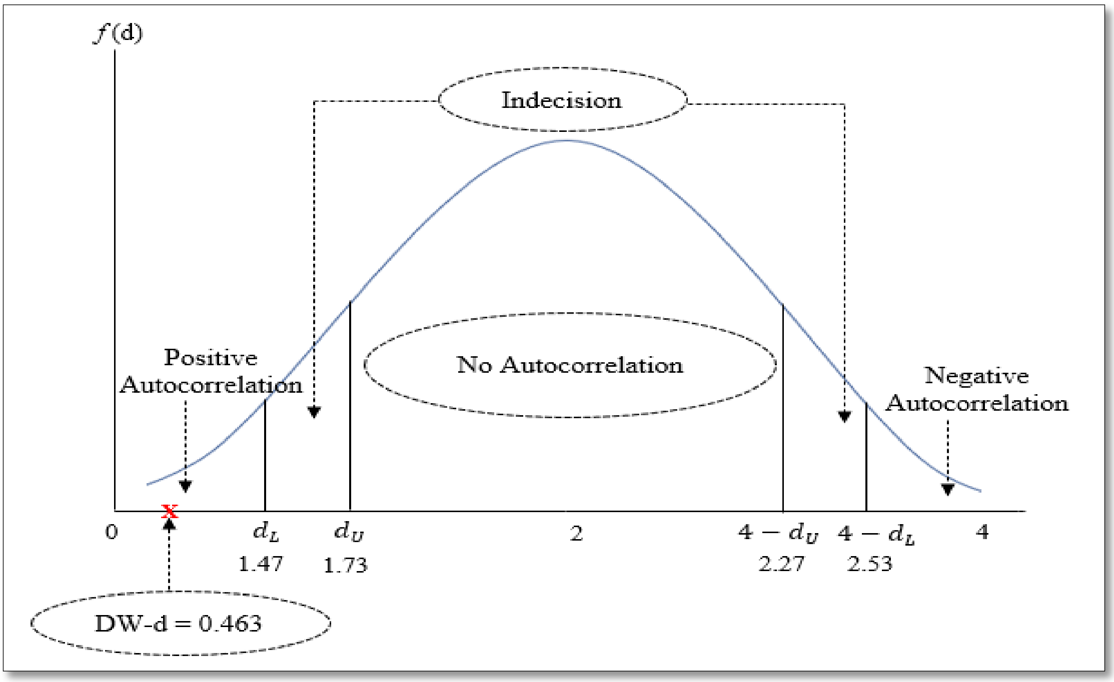

Figure 4: Result of Durbin-Watson d statistic

Yule-Walker estimation is used in this case, it is a method that used to correct autocorrelation with autoregressive of order one [AR(1)] errors as well as exogenous independent variables. The transformed Durbin-Watson d statistic is 1.9686 and it is larger than the value of d_U (1.73). Thus, the problem of autocorrelation is resolved.

Artificial Neural Network

This study will only focus on multi-layer perceptron (MLP) neural network which also a typical example of feed-forward neural network with Levenberg Marquardt Algorithm (LMA). LMA is used due to being the fastest supervised algorithm for training and widely used for time series prediction in the ANN model (Beale et al., 2017). By taking the quarterly data of PHPI from 2005Q1 to 2022Q1[P], the data was forecasted from 2022Q2 to 2023Q4. The forecasting techniques were implemented by using Matlab©R2019a software under neural networks toolbox to create, train, test and validate the ANN model.

Creating and Training the ANN

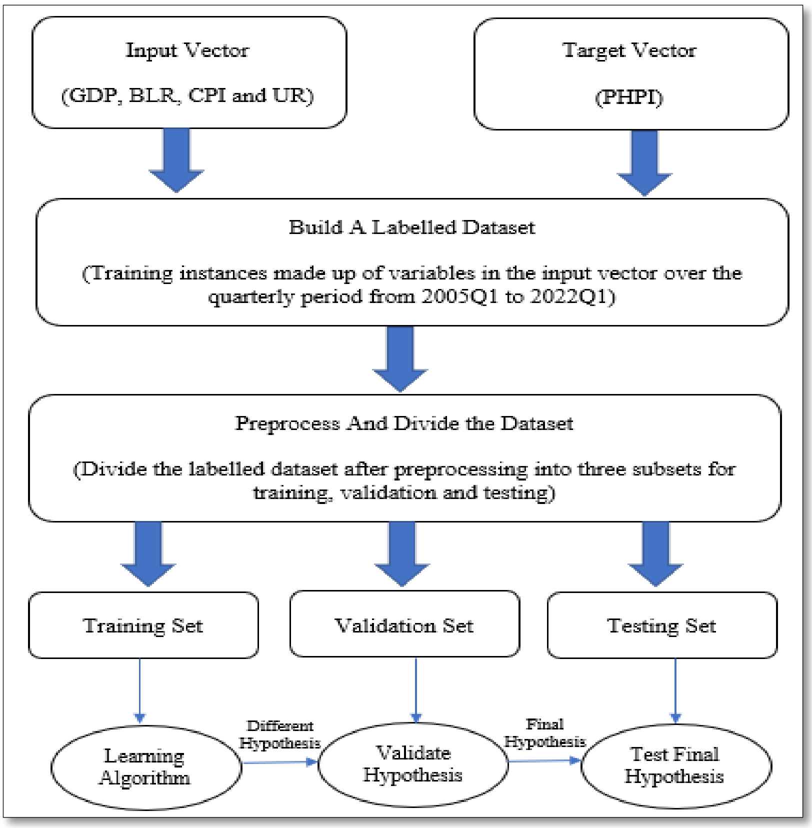

Figure 5: Collection and distribution of dataset

Collecting and Preparing Data

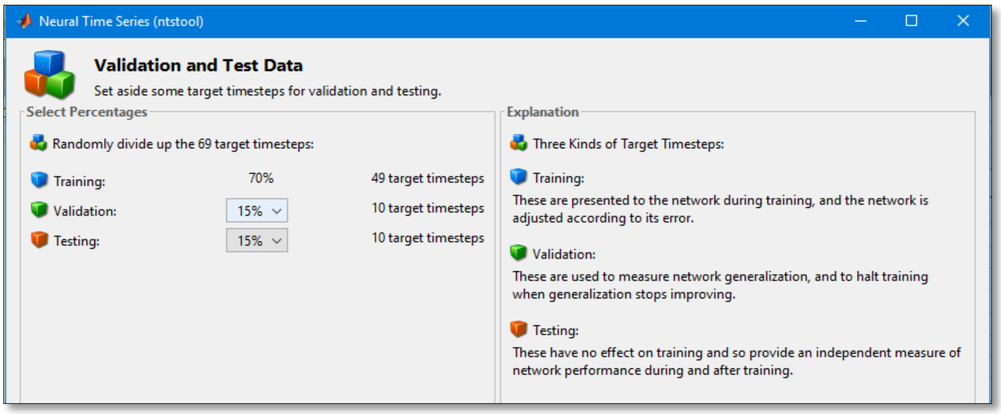

Figure 6: Validation and test data using Matlab©R2019a

Three subsets are used to train multilayer neural networks. First subset is the training set is used to compute the gradient and update the network weight and biases. The second subset is called the validation set and it is used for validating the set. The testing set is the third subset, and it is used to test the set-in order to check for test errors in the network. Figure 6 shows the driver and function in Matlab©R2019a that used to divide input vectors and target vectors into three sets as follows:

- 70% will be used for training
- 15% will be used to validate that the network is generalizing and to stop training before overfitting.
- The last 15% will be used as a completely independent test of network generalization.

Training and Testing the Network

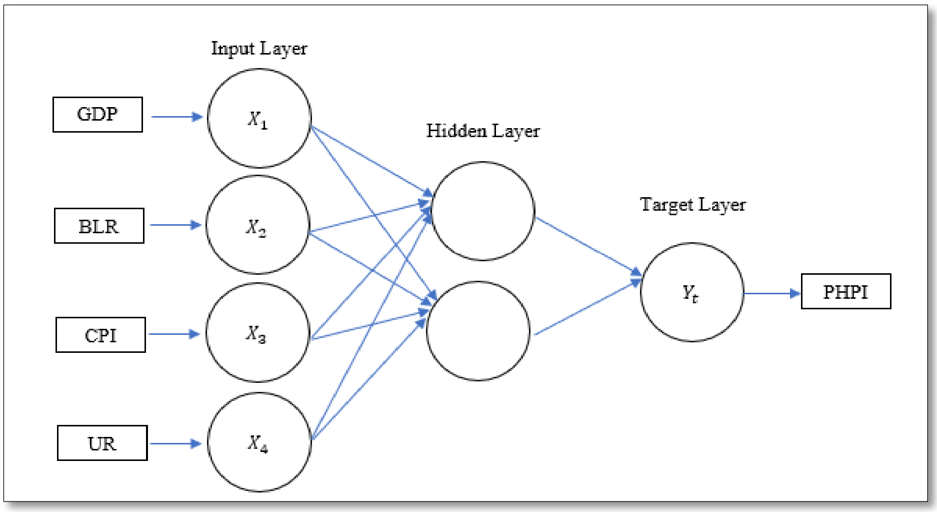

Figure 7: ANN structure of the proposed model for PHPI.

After preparing the dataset, a feed-forward neural network is created and configured with 4 input nodes, 2 hidden layers with 10 neurons, and 1 output neuron as shown in Figure 7 which illustrates the ANN structure of the proposed model for PHPI.

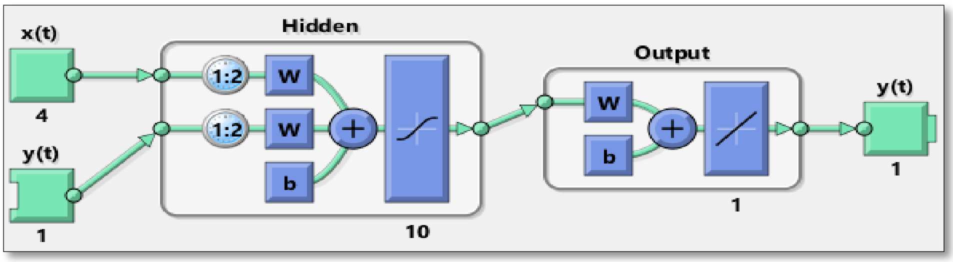

Figure 8: Open Loop design in NARX neural network architecture for PHPI 4-10-1 using Matlab version 2019a

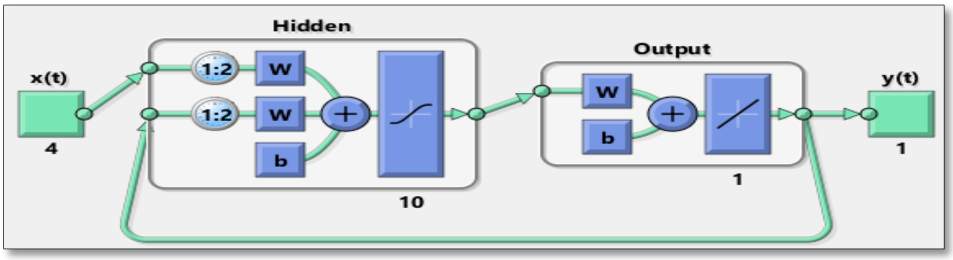

Figure 9: Close Loop design in NARX neural network architecture for PHPI 4-10-1 using Matlab version 2019a

Nonlinear Autoregressive Network with Exogenous (external) Input, or NARX was adopted in this study because it has proven to be the most effective and accurate solution for multivariable data series to predict future value of a time series based on the past values (Abraham et al., 2020) (Beale et al., 2017). The open and close loop design in NARX architecture model is illustrated in Figure 8 and Figure 9 by using the neural network toolbox of Matlab©R2019a software to create the model for the chosen neural network of PHPI 4-10-1 which contains of 4 neurons in the inputs layer, two hidden layers with 10 neurons in it and 1 neuron in the output layer.

When the loop is open on the NARX network as shown in Figure 8, it is performing a one-step ahead prediction. Meanwhile, with the loop closed as shown in Figure 9, it is used to perform multi-step-ahead predictions (Perez, 2019). The whole training process including the validation and testing processes is carried out in open loop. In a typical workflow, the network is first completely constructed in open loop, and only after it has been trained (which includes steps like validation and testing) is it converted to closed loop for multistep-ahead prediction. This is because training the network requires validation and testing steps. In this study, the number of hidden neurons and number of delays is set at default and not adjusted as shown Figure 8. The network was trained with the well-known Levenberg-Marquardt algorithm (LMA) which is commonly used as *trainlm* in Matlab©R2019a.

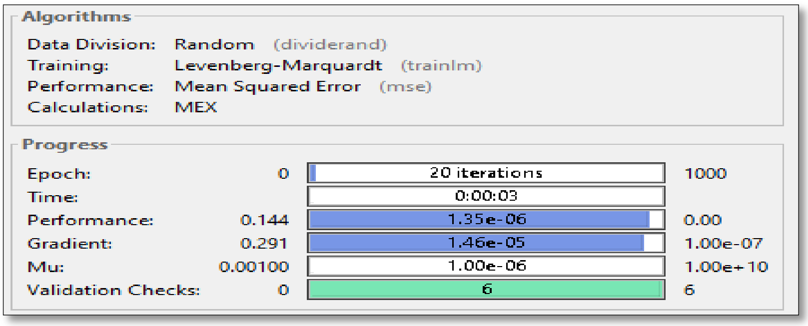

Figure 10: Training progress of the proposed model for PHPI

PHPI training reached an optimal value for the regression and correlation among the variables after 20 iterations as shown in Figure 10. The training procedure stops when the performance on the test data does not improve following a fixed number of training iterations which indicates that the network has reached the accuracy (MathWorks, 2022). As shown in Figure 10, the training continued until the validation error failed to decrease for six iterations (validation stops).

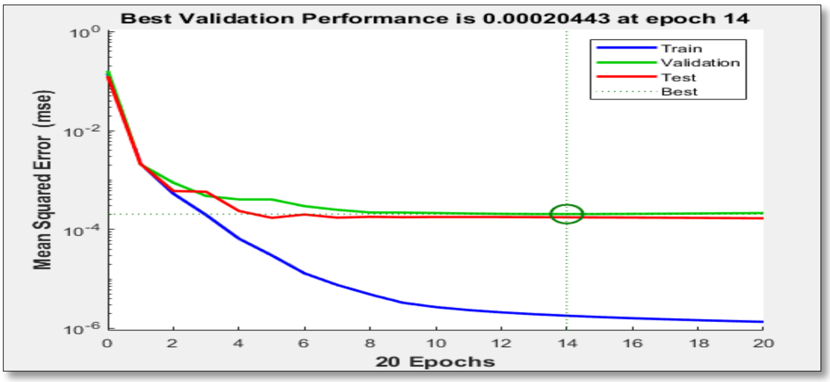

Figure 11: Training performance of the proposed model for PHPI

Meanwhile, once the input data had been trained, then its performance was examined using means squared error (MSE) which is in log scale as shown in Figure 11. It shows that the best validation performance is achieved at epoch 14 with very small final mean squared error (MSE) value at 0.00020443.

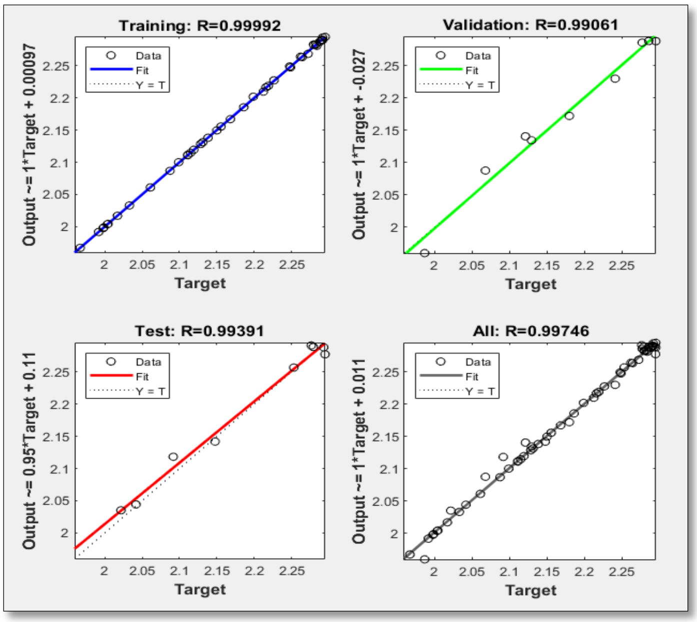

Figure 12: Regression for training, validation, testing and all data of the PHPI

The regression of this study is shown in Figure 12. The three axes represent the training, validation, and the testing data. The dashed line in each axis represents the optimum result (outputs = targets). The solid line represents the best fit linear regression line between outputs and targets. The *R* value is an indication of the relationship between the outputs and targets. Therefore, Figure 12 shows the output of regression and correlations of the PHPI model, the fit was well aligned which indicates the model has a good capacity for generalization and prediction. PHPI training reached an optimal value for the regression and correlation among the variables after 20 iterations and pose an *R* value 0.99746 which is near to 1.

Validation of the PHPI data

Table 6: Actual and Forecasted data on PHPI

Time Series	Actual Data	Forecasted Data	[Actual - Forecasted]	Absolute Percentage Error (%)
2021Q1	194.2	194.00	0.20	0.10
2021Q2	196.4	192.70	3.70	1.88
2021Q3	190.2	192.60	0.01	1.26
2021Q4	191.6	190.00	1.60	0.84
2022Q1	193.5	189.00	4.50	2.33
			MAPE(%)	1.28

PHPI has been forecasted for 5 quarters ahead using the ANN-NARX approach based on the actual data of year 2021Q1 to 2022Q1. Table 6 shows the forecasting performances that were evaluated by computing the MAPE between actual and predicted values of PHPI. Based on

Table 6, the evaluation using MAPE value of 1.28 shows that the Neural Network is capable and can be classified as very good to forecast PHPI. The findings indicates that the ANN model with the developed structure should be able to produce a good prediction with least amount of error and finally this neural network could be a useful tool for forecasting PHPI.

Comparisons of Error Performance Measurement

Forecasting of PHPI models are then validated using some other indicators, indicator used in this study are MSE, RMSE, MAPE and MABE.

Table 7: Comparison of error performance measurement on ANN-NARX model for PHPI

Error Measures	ANN-NARX model for PHPI
MSE	29.8223
RMSE	5.4601
MAPE	2.4154
MABE	3.1824

Based on Table 7, all the statistical test results indicated good fit of PHPI model with low values of MSE, RMSE, MAPE and MABE. Meanwhile, MAPE showed a percentage error of 2.41% which is less than 10% error. This implies that the ANN model is able to predict PHPI with low errors.

PHPI Quarterly Forecast 2022Q2 to 2023Q4

A graph is plotted between the actual and predicted outputs data of PHPI produced by using Matlab©R2019a. Plotting a graph is the best method to compare the actual data and predicted output of PHPI. The accuracy of the model can be analysed by using the graph as shown in Figure 13.

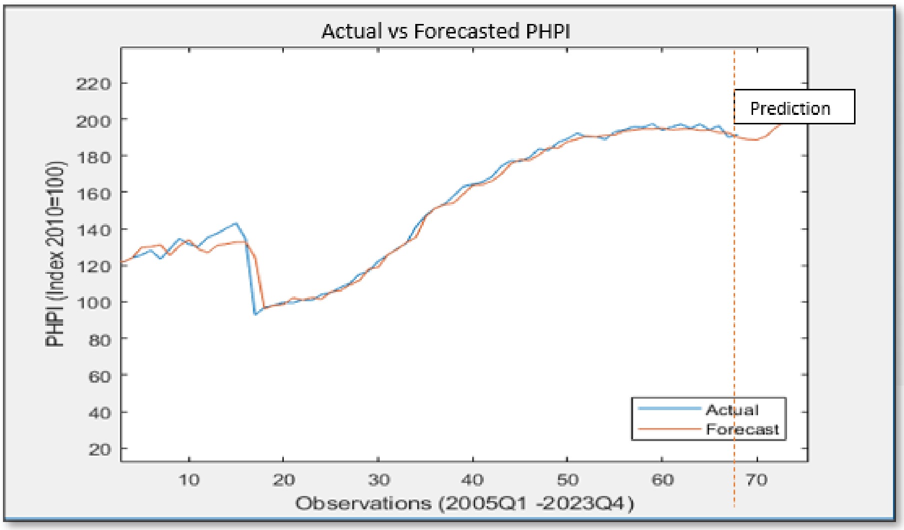

Figure 13: Multistep prediction of PHPI using Matlab version 2019a

The blue line (target) represents the original data whereas the red line (forecast) represents the obtained values for each quarter. PHPI data has been forecasted from year 2022Q2 to 2023Q4 using ANN with the Nonlinear Autoregressive Network with Exogenous Input (NARX) approach along with the graphical visualization as shown in Figure 13.

Table 8: Results of forecasted data of PHPI from 2022Q2 to 2023Q4

Forecasted PHPI Data							
Observations	70	71	72	73	74	75	76
Quarterly per year	2022Q2	2022Q3	2022Q4	2023Q1	2023Q2	2023Q3	2023Q4
PHPI	188.76	190.76	195.48	198.73	200.67	201.83	203.04

Moreover, Table 8 shows the results of multistep ahead forecasted data of PHPI from 2022Q2 to 2023Q4 which is generated in close-loop form of ANN-NARX. Other than that, ANN-NARX predictions has shown a better adjustment compared to the original data as it showed a smoother follow-up. The multi-layer perceptron (MLP) network of PHPI has shown a good performance and reasonable prediction accuracy that was achieved in this model as the actual and predicted PHPI values are very close to each other which is clearly establish the reliability and efficiency of the proposed model.

5. Conclusion

Based on the summary of findings from previous chapter, the objectives have been successfully achieved. The unit root test is necessary as this research employed time series data. Although the selected variables in level form contain unit (non-stationary), however, they were proved to be stationary through first difference transformation. It means that *ln*GDP, BLR, *ln*CPI and UR are stationary in first difference. Besides, the model's goodness of fit recorded at 0.7858. It means that 78.58% of changes in *ln*PHPI can be explained by the selected independent variables, while the remaining 21.42% can be explained by the error term. The goodness of fit might able to improve if quarterly data sets such as household income and expenditure, construction cost and real property gain tax are available.

Results from the OLS estimation suggested that variables such *ln*GDP, BLR and UR have a significant impact at 1% significant level whereas *ln*CPI is insignificant at 1%, 5% and 10% significance level on the PHPI in Malaysia and their signs are correct accordingly. Although lnGDP, BLR and UR are statistically significant at 1% significant level, but most important determinant of PHPI in Malaysia is *ln*GDP. This is because *ln*GDP recorded a relatively larger magnitude of impact on PHPI as compared to other variables. It is fairly reasonable to conclude that variables including *ln*GDP, BLR and UR have a significant impact on PHPI in Malaysia. These variables ought to be given attention as they play roles in affecting PHPI in Malaysia. Moreover, *ln*CPI which is insignificant need strong evidence to provide support the effect exists in the model. Next, multicollinearity and autocorrelation test were conducted. The model is free from multicollinearity problem as all the computed VIF were less than 10. In addition, Durbin-Watson d test revealed that the model suffers from positive autocorrelation at first. Nevertheless, the autocorrelation problem was then successfully resolved through Yule-Walker estimation.

Meanwhile, the ANN with the Nonlinear Autoregressive Network with Exogenous Input (NARX) has used in this study to generate a forecasting model for PHPI. The analysis shows that the best ANN-NARX approach to forecast PHPI in Malaysia is 4-10-1 which contains of 4 neurons in the inputs layer, two hidden layers with 10 neurons in it and 1 neuron in the output layer. This method also aims to help decision makers to analyze the market structure and compare the recent HPI with their findings. In the model, 4 different inputs are used, and the target of the system is the future PHPI. Buying a house or using a mortgage is a very difficult opinion for most of the people. A proper HPI forecast can be very helpful for future references. Thus, the proposed method can be utilized to ease the decision-making process particularly for both property investors and final consumers.

References

Abraham, E. R., Mendes dos Reis, J. G., Vendrametto, O., Oliveira Costa Neto, P. L. D., Carlo Toloi, R., Souza, A. E. D., & Oliveira Morais, M. D. (2020). Time series prediction with artificial neural networks: An analysis using Brazilian soybean production. *Agriculture*, *10*(10), 475.

Beale, M. H., Hagan, M. T., & Demuth, H. B. (2017). *Neural Network Toolbox User's Guide.* Natick, MA: The MathWorks, Inc. Retrieved from www.mathworks.com

Construction Plus Asia. (2022). Construction Forecast and Review 2H 2021 & 1H 2022: Malaysia. Retrieved from https://www.constructionplusasia.com/my/construction-forecast-and-review-2h-2021-1h2022-malaysia/

Forbes. (2018). Six fundamental human needs we need to meet to live our best lives. Retrieved from https://www.forbes.com/sites/quora/2018/02/05/six-fundamental-human-needs-we-need-to-meet-to-live-our-best-lives/?sh=76e8434344a

Garg, A. (2016). Statistical methods for estimating house price index. *Journal of Business & Financial Affairs*, *5*(4), 1-3.

Ge, B., Ishaku, M. M., & Lewu, H. I. (2021). Research on the effect of artificial intelligence real estate forecasting using multiple regression analysis and artificial neural network: A case study of Ghana. *Journal of Computer and Communications*, *9*(10), 1-14.

Geerolf, F., & Grjebine, T. (2014). Assessing house price effects on unemployment dynamics. *CEPII Working Papers*. Retrieved from https://econpapers.repec.org/paper/ciicepidt/2014-25.htm

Gregor, A. (2020). *The New York Times*. House hunting in Malaysia: A restored rowhouse in Penang. Retrieved from https://www.nytimes.com/2020/04/15/realestate/house-hunting-in-malaysia-a-restored-rowhouse-in-penang.html

Jehani, N. A., Mastani, N. A., Saudin, S., & Ab Malek, I. (2020). A study on the relationship between house price index and its determinants in Malaysia. *Malaysian Journal of Computing (MJoC)*, *5*(2), 515-522.

Kitapci, O., Tosun, O., Tuna, M. F., & Turk, T. (2017). The use of Artiicial Neural Networks (ANN) in forecasting housing prices in Ankara, Turkey. *Journal of Marketing and Consumer Behaviour in Emerging Markets, 1*(5), 4-14. Retrieved from https://ideas.repec.org/a/sgm/jmcbem/v1i5y2017p4-14.html

Latif, N. S. A., Rizwan, K. M., Rozzani, N., & Saleh, S. K. (2020). Factors affecting housing prices in Malaysia: A literature review. *International Journal of Asian Social Science*, *10*(1), 63-67.

Lee, Y., & Azlan, M. I. B. (2022). Determinants of housing prices: Evidence from Malaysia and Singapore. *Asian Journal of Research in Education and Social Sciences*, *3*(4), 91-106.

MathWorks. (2022). What is a neural network? 3 things you need to know. Retrieved from https://ww2.mathworks.cn/en/discovery/neural-network.html

Mohan, S., Hutson, A., MacDonald, I., & Lin, C. C. (2019). Impact of macroeconomic indicators on housing prices. *International Journal of Housing Market and Analysis, 12*(6), 1055-1071.

NAPIC. (2021). Malaysian House Price Index Q1-Q3 2021. *National Property Information Centre*. Retrieved from https://napic.jpph.gov.my/portal/web/guest/publication

Narkhede, S. (2018). Understanding descriptive statistics. *Towards Data Science*. Retrieved from https://towardsdatascience.com/understanding-descriptive-statistics-c9c2b0641291

Perez, C. (2019). *Big Data Analytics With Neural Networks Using MATLAB.* Texas: Lulu.com. Retrieved from https://books.google.com.my/books?id=6PPGDwAAQBAJ&pg=PT188&lpg=PT188&dq=multistep+prediction+

for+artificial+neural+network+using+matlab&source=bl&ots=X8Rt6_lzw2&sig=ACfU3U2reJYjVtvv-ZriA00n3IsutGq8Pw&hl=en&sa=X&ved=2ahUKEwj33-in-ar5AhUd-TgGHQjfAfg4PBDoAXo

Radzi, M., & MS, M. C., Kamarudin, N., & Mohammad, IS (2012). Forecasting house price index using artificial neural network. *International Journal of Real Estate Studies*, *7*(1), 43-48.

Rahman, M. F., & Ridzuan, A. R. (2020). Factors that determine house price index in Malaysia. *International Journal of Academic Research in Accounting, Finance and Management Sciences, 10*(1), 351-359. doi:10.6007/IJARAFMS/v10-i1/7224

Sukrri, N. N., Wahab, N. A., & Yusof, R. M. (2019). An enhanced house price index model in Malaysia: A Laspeyres approach. *International Journal of Economics, Management and Accounting, 27*(2), 373-396. Retrieved from https://journals.iium.edu.my/enmjournal/index.php/enmj/article/view/631

Sun, N. (2021). The effects of housing price on unemployment rate and stock market. *International Journal of Trade, Economics and Finance, 12*(5). doi:10.18178/ijtef.2021.12.5.707

Tan, Y. (2011). A hedonic model for house price in Malaysia. *International Real Estate Conference*, (pp. 12-15). Retrieved from http://www.prres.net/papers/tan_an_hedonic_model_for_house_prices_in_malaysia.pdf

Trofimov, I. D., Aris, N. M., & Xuan, D. C. (2018). Macroeconomic and demographic determinants of residential property prices in Malaysia. *Munich Personal RePEc Archive*, 1-24. Retrieved from https://mpra.ub.uni-muenchen.de/85819/1/MPRA_paper_85819.pdf

Tumbarello, P., & Wang, S. (2010). What drives house prices in Australia? A+L4584 Cross-Country Approach. *2010 International Monetary Fund*, *10*(291), 1-22. doi:10.5089/9781455211722.001