J. Ris. & Ap. Mat. Vol. 08 No. 02 (2024) pp. 94-104
Jurnal Riset dan Aplikasi Matematika (JRAM)
e-ISSN: 2581-0154
URL: journal.unesa.ac.id/index.php/jram

THE COMPARISON OF HEDONIC REGRESSION AND ARTIFICIAL NEURAL NETWORK IN THE DEVELOPMENT OF MASS APPRAISAL MODEL

Case Study: Residential Property in Surabaya City

MUKTI RATNA DEWI[1*], BRODJOL SUTIJO SUPRIH ULAMA[2], DESTRI SUSILANINGRUM[3]

[1, 2, 3] Institut Teknologi Sepuluh Nopember

*mukti_ratna@its.ac.id

ABSTRACT

The appraisal of land and building tax objects in Surabaya City has been carried out manually, so it is prone to bias, human error, and inefficiency in terms of time and energy. Therefore, it is necessary to develop a mass appraisal system based on a property price prediction model in accordance with the real estate market structure. The traditional approach to mass appraisal is based on multiple regression analysis, which, despite having a long history in the appraisal world, cannot address interactions between variables such as nonlinearity. This deficiency was then overcome by introducing an Artificial Intelligence approach based on an artificial neural network (ANN), which has the ability to learn on its own and generalize solutions. Specifically, this study aims to build a mass appraisal model for residential property using (i) multiple regression analysis and (ii) artificial neural network, as well as (iii) determining the best mass appraisal model. The result of the study shows that ANN performs better than multiple regression analysis in predicting the value of residential property in Surabaya City.

Keywords: Artificial Neural Network, Multiple Regression Analysis, Mass Appraisal.

1. Introduction

Tax in Indonesia is the primary source of state revenue besides non-tax and grants, which are divided into seven sectors, one of which is the Land and Building Tax (PBB). PBB included in property tax uses Taxable Object Sales Value (NJOP) as the basis for taxation. PBB tax objects can be assessed using several approaches, one of which is the market comparison approach. In this approach, the market value of a property is based on the selling price of other comparable properties [1]. In practice, a property to be assessed will only be compared with two or three properties with the same characteristics in the vicinity. For tax assessment purposes involving many properties, this manual assessment process is time-consuming, costly, inefficient, and prone to bias and human error. Therefore, it is necessary to develop a mass assessment method that is under the structure of the real estate market and is adaptive to changes over time.

Based on KEPI & SPI issued by the Indonesian Appraisers Professional Society (MAPPI), a mass appraisal is defined as a systematic appraisal system for a group of individual properties/assets based on existing data and using standard procedures and statistically tested [2]. A model for estimating the property's value needs to be developed in mass appraisal. This model is usually based on a comparative approach that assumes the property's market value can be determined based on other properties whose selling and buying prices and influencing factors

2020 *Mathematics Subject Classification*: 62M45
Diterima: 24-09-24; direvisi: 17-10-24; diterima: 21-10-24

are known [3]. Meanwhile, mass appraisal with the help of a computer or Computer Assisted Mass Appraisal (CAMA) is a property appraisal system, usually only a few types of tangible property, that combines computers supported by statistical analysis such as multiple regression analysis and adaptive estimation procedures to assist appraisers in estimating value [2].

The traditional approach to mass appraisal is based on Hedonic Regression, better known as Multiple Regression Analysis (MRA). This method is popular because of its good methodology and long application history. It is also widely accepted among practitioners and academics. In Indonesia, the development of a property appraisal model using hedonic regression has been applied to the appraisal of apartments [4], [5], houses [6]–[8], rice fields [9], [10], and land [10]. Although property appraisal using multiple regression has been widely used, problems such as multicollinearity, interaction between independent variables, heteroscedasticity, nonlinearity and outliers can seriously affect the performance of hedonic models in real estate appraisal [11], [12]. To overcome these problems, several Artificial Intelligence (AI)-based methods need to be introduced into mass property appraisal research. The most frequently studied AI method is Artificial Neural Network (ANN). ANN is an artificial intelligence (AI) model initially designed to replicate the human brain's learning process. It is considered capable of overcoming problems in MRA because it has the ability to learn on its own, generalize solutions, and respond to highly correlated, incomplete or previously unknown data. Several studies have reported that ANN-based approaches produce better results than MRA [13], [14].

Based on discussions with the Head of the East Java DPD MAPPI (Indonesian Professional Appraisers Society), it was obtained that the mass appraisal model for properties in Surabaya is still unavailable, resulting in the appraisal and justification of taxation that cannot be verified. Therefore, in this study, a case study of the development of mass appraisal will be carried out in Surabaya. This study limits the scope of the study in the Surabaya area, with the definition of property as real estate, which only includes private residences as samples.

2. Literatur Review

2.1 Hedonic Regression

The hedonic price theory assumes that a commodity, such as a house, can be viewed as an aggregate of individual components or attributes. Consumers are assumed to purchase goods with a set of attributes that maximize the underlying utility function. The advantage of the hedonic model is that it can estimate value based on actual attributes. In this model, all attributes that affect the value of a property are analyzed together, and the level of influence of each attribute can also be studied. Research conducted by Fahirah [15] and Supriyono [16] states that several attributes that affect the value of a property include location, land area, building area, number of bathrooms, number of bedrooms, environment, and the presence of a security system.

Hedonic regression is a hedonic model that uses MRA. MRA has been used as the main mass real estate appraisal method, which has the general equation (1).

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \varepsilon \tag{1}$$

where:

- Y is the response or house prices in this study;
- $X_1, \ldots, X_p$ is a predictor or a factor that influences house prices;
- $\beta_0, \ldots, \beta_p$ is a parameter coefficient that measures the change in the response associated with a one-unit change in the predictor when all other predictors are held constant;
- ε is the residual in which ε ~ IIDN$(0,\sigma^2)$.

In Indonesia, the development of property appraisal models using hedonic regression has been applied to the appraisal of apartments [4], [5], houses [6]–[8], rice fields [9], [10], and land

[10]. Although property appraisal using MRA has been widely used, problems such as multicollinearity, interaction of independent variables, heteroscedasticity, nonlinearity and outliers can seriously affect the performance of hedonic models in real estate appraisal [12]. To overcome this problem, several Artificial Intelligence (AI)-based methods need to be introduced into mass property appraisal research.

2.2 Artificial Neural Network

Artificial Neural Network (ANN) is a branch of Artificial Intelligence (AI) often developed in regression and classification modelling. ANN is a form of AI that attempts to imitate the human brain and nervous system. The structure of ANN usually consists of a series of processing elements (PE) or nodes arranged in an input layer (in this case, the property price and its attributes), a hidden layer (black box), and an output layer (price prediction) as shown in Figure 1.

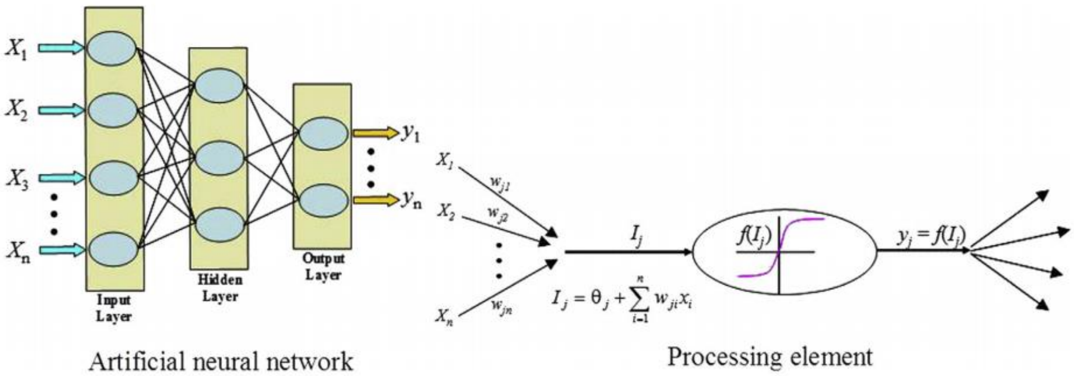


Figure 1. Typical Structure and Operation of ANN [17]

The input of each PE in the layer x_i is multiplied by an adjustable connection weight w_{ij}. At each PE, the weighted input signals are summed and a threshold value θ_j is added. The combined inputs I_j are then passed through a nonlinear transfer function $f(.)$ to produce the output of PE y_j. The output of one PE provides the input to the PE in the next layer. This process is summarized in equations (2) and (3) and illustrated in Figure 1.

$$I_j = \sum_{j=1}^{p} w_{ji}x_i + \theta_i \qquad \text{combined} \tag{2}$$

$$y_j = f(I_j) \qquad \text{transfer} \tag{3}$$

ANN is one of the branches of AI that is widely developed in mass assessment [13], [14], [18]–[20]. Several studies have reported that ANN-based approaches produce better results compared to MRA [13], [14], [19].

2.3 Model Performance Evaluation

The residential property assessment model formed using hedonic regression and ANN methods was evaluated using error metrics in the form of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) with the formulas given in equations (4), (5), and (6) respectively.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \quad (5)$$

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{y_i - \widehat{y}_i}{y_i}\right) \times 100\% \quad (6)$$

where:
- n is the amount of data.
- y_i is the actual value with $i = 1, ..., n$.
- $\widehat{y}_i$ is the estimated value with $i = 1, ..., n$.

3. Research Methodology

3.1 Data Description

The data for this study were collected from property advertisement websites, which were then re-verified with the seller to obtain a price estimate close to reality, considering that price data from websites tends to be high. The research variables used follow the data variables from property advertisements, as shown in Table 1. The target population is all residential houses in Surabaya. To ensure that the data represents the condition of the Surabaya property market, each sub-district in Surabaya is represented by a minimum of 5 houses. Samples are taken using purposive sampling according to the criteria: residential properties transacted or offered in the past year. Data was collected for approximately three months, from July - September 2023, obtaining 548 samples.

Table 1. Research Variables

No.	Variable	Description
1	House price	House selling price (million)
2	Land area	The total of the land the property is on (m^2)
3	Built-up area	The size of the property where space can be used (m^2)
4	Baths	Number of bathrooms
5	Beds	Number of bedrooms
6	Story	Number of stories
7	Rights	Types of property rights (1: Freehold Title – SHM; 2: Building Rights Title – HGB; 3: Others)

3.2 Research Framework

The steps of the analysis in this study are as follows:

- Step 1: Exploratory analysis of residential property selling price data and the factors that influence it.
- Step 2: Divide the data into training data and testing data.
- Step 3: Building a mass assessment model using hedonic regression on training data.
  - (i) Create dummy variables for ownership status variables.
  - (ii) Estimate multiple regression model parameters.
  - (iii) Conduct partial and simultaneous tests and see adjusted R^2.
  - (iv) Conduct assumption test $\varepsilon \sim IIDN(0, \sigma^2)$.
- Step 4: Building a mass scoring model using ANN on training data.
  - (i) Encode the ownership status variable and rescale the data.
  - (ii) Determine the number of hidden layers.

(iii) Determine the number of neurons in each layer.
(iv) Run the backpropagation algorithm.
(v) Calculate the relative contribution factor.

- Step 5: Determine the best mass appraisal model between MRA and ANN by comparing the performance of both models.

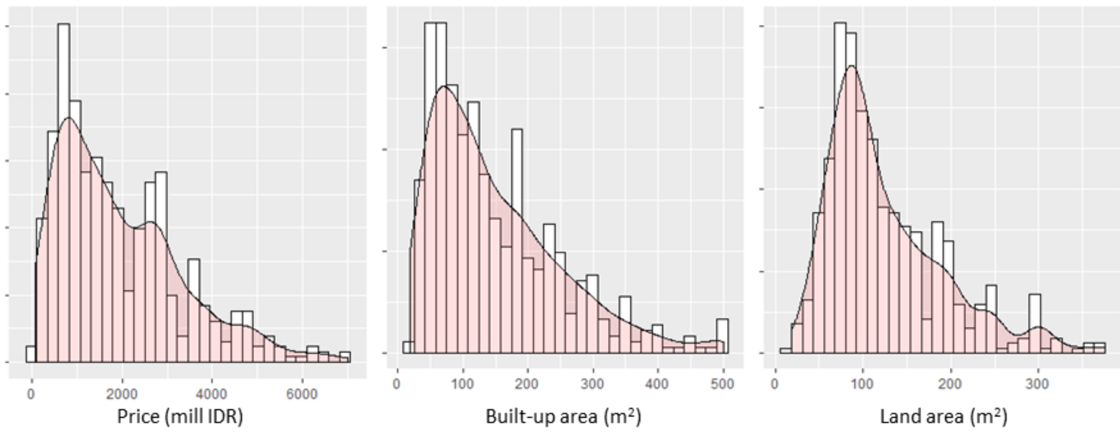
4. Results and Discussion

4.1 Data Exploration

Based on Table 2, the selling price of residential property in Surabaya City has an average of 1.935 billion rupiah, with the lowest and highest selling prices being 75 million and 7 billion rupiah, respectively. The distribution of this selling price data is skewed to the right, as shown in Figure 2 (a). In line with the selling price data, the distribution of building area and land area data in Figures 2 (b) and (c) is also skewed to the right, with an average of 149.85 m^2 and 125.30 m^2, respectively. Meanwhile, most of the houses sold have two floors with facilities of three bedrooms and one bathroom, as presented in Table 2.

Table 2. Summary Statistics

Variable	Mean	Variance	Min.	Median	Max	Modus
House price	1,935.80	1,920,533.70	75	1,600	7,000	2,900
Land area	149.85	10,230.72	18	120	500	60
Built-up area	125.30	4,482.82	18	105	375	90
Beds	3.35	1.25	1	3	6	3
Baths	2.52	1.65	1	2	6	1
Story	1.66	0.25	1	2	3	2



(a) (b) (c)

Figure 2. The Distribution of (a) House Selling Price, (b) Built-up Area, and (c) Land Area

As seen in Figure 3, the house selling price with building area and land area has a positive linear relationship with correlation values of 0.91 and 0.87, respectively. Meanwhile, the average selling price of a house also appears to increase along with the increasing number of bedrooms, bathrooms, and floors, with correlation values of 0.73, 0.79 and 0.54, respectively. On the other hand, from the collected residential property data, 71% have SHM certificate type, 20% have HGB certificate, and 9% have other types of legal certificates (PPJB, Girik, Adat, and others). Based on Figure 4, the house selling price with HGB certificates has the largest median compared to other ownership certificates.

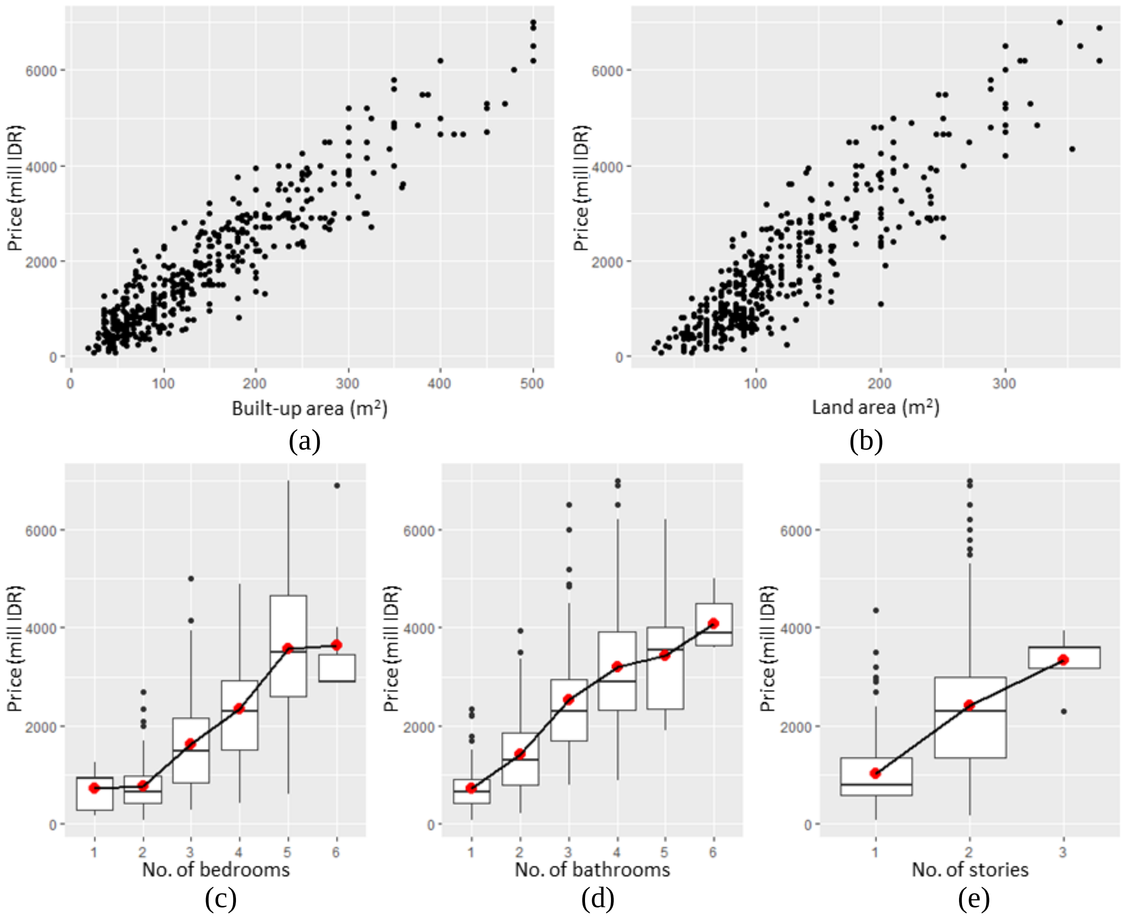


Figure 3. Relationship Between House Selling Price and (a) Building Area, (b) Land Area, (c) Number of Bedrooms, (d) Number of Bathrooms, and (e) Number of Stories.

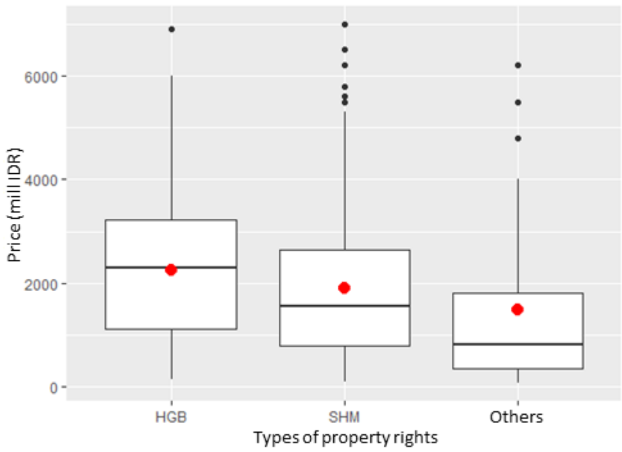


Figure 4. Comparison of House Selling Prices Between Types of Ownership Certificates

4.2 Hedonic Regression Model

The results of the data exploration analysis show that there is a linear relationship between the selling price of a house and the building area, land area, number of bedrooms, number of bathrooms, and number of floors, so these six variables will be included as predictors in the multiple linear regression analysis. Meanwhile, the ownership status variable is also included in the regression model estimation even though the results of the data exploration analysis show that the type of ownership certificate tends not to affect the selling price of a house.

Before running the regression analysis, the data was partitioned into two parts, namely training and testing, with a ratio of 80:20. The training data was then used to estimate the multiple linear regression model parameter. The results are presented in Table 3.

Table 3. Model Estimation of Residential House Selling Price

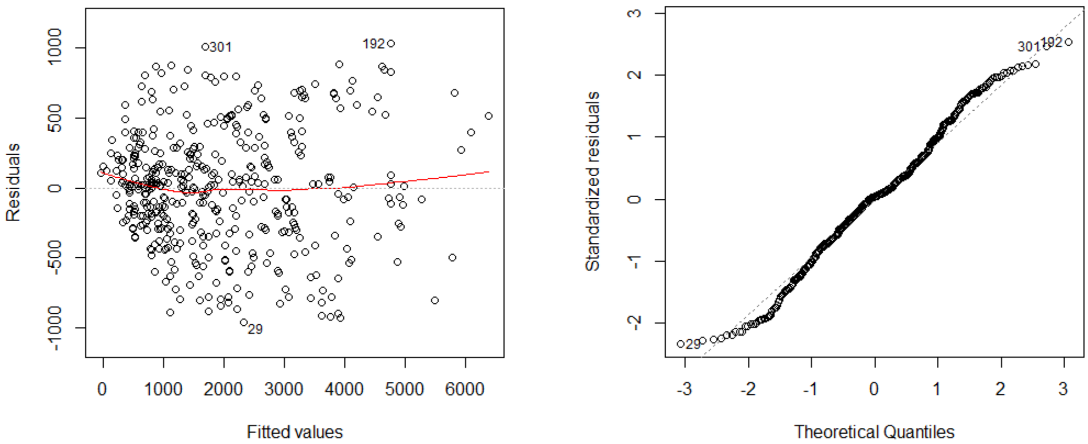
Variable	Coef.	SE Coef.	*t-value*	*p-value*	VIF
Intercept	-554.09	94.47	-5.86	<0.05	7.11
Built-up area	6.32	0.52	12.22	<0.05	5.36
Land area	8.36	0.67	12.37	<0.05	3.30
Beds	-110.12	31.15	-3.54	<0.05	3.55
Baths	218.43	27.89	7.83	<0.05	1.84
Story	188.99	51.79	3.65	<0.05	1.84
Rights – HGB	54.72	49.92	1.09	0.27	1.11
Rights – other than HGB	-16.39	70.56	-0.23	0.82	
RSE: 411.60 with df: 461. R-sq: 0.91 and Adj. R-Sq: 0.91. F-stat: 655.7 (df_1: 7; df_2: 471) and p-value: <0.05.					

A mass appraisal estimation model for residential properties based on the estimation results in Table 3 is as follows:

$$\text{House price} = -554{,}09 + 6{,}32 \text{ built-up area} + 8{,}36 \text{ land area} - 110{,}12 \text{ beds} + 218{,}43 \text{ baths} + 188{,}99 \text{ story} + 54{,}72 \text{ HGB right} - 16{,}39 \text{ other rights} + \varepsilon \quad (7)$$

The regression model in equation (7) has an adjusted R^2 of 91%, and all predictors are significant at the 5% level except for the type of ownership certificate. However, this variable is maintained in the model considering that in the process of buying and selling a house, the legality aspect is one of the factors affecting the buyer's decision.

The estimation results in Table 3 show that the building area, land area, the number of bathrooms and stories, as well as HGB ownership certificate positively affect a house's selling price. In contrast, the number of bedrooms and ownership certificates other than HGB and SHM can reduce the value of a house. In practice, the number of bedrooms does not always increase the house's selling price because too many bedrooms can reduce the availability of space that can be used for other things. It can also indicate a nonlinear relationship between the selling price of a house and the number of bedrooms that cannot be captured by multiple linear regression.



(a) (b)

Figure 5. Residual Assumptions of (a) Equal Variance and (b) Normality

In general, houses with larger land and buildings and more bathrooms and floors carry more value. The average selling price of a house with an HGB certificate is 54.72 million higher than a house with an SHM certificate. In contrast, houses with other certificates have an average selling price of 16.39 million lower than houses with an SHM certificate.

When checking the residual assumption, a fairly high VIF value was obtained for the building area and land area variables, as presented in Table 3. However, multicollinearity between these predictors can still be tolerated because VIF <10. The identical assumption can be evaluated through Figure 5 (a), where the regression model residuals do not form a particular pattern and are evenly spread on both sides of the zero line. Figure 5 (b) shows the model meets the normality residual assumption.

4.3 Back Propagation ANN

ANN is one of the techniques in machine learning that requires input variables (predictors) and output (responses) in the form of numbers (numeric), so dummy variables need to be created for the ownership variables first. Afterward, the data is rescaled so all variables are on a scale of 0 to 1. Theoretically, this stage is not mandatory since the activation function will naturally carry out the rescaling process. However, in practice, non-rescaled data can cause the ANN algorithm to fail to converge before reaching maximum iteration. The rescaled data is then divided into training and testing data with the same composition as multiple linear regression analysis.

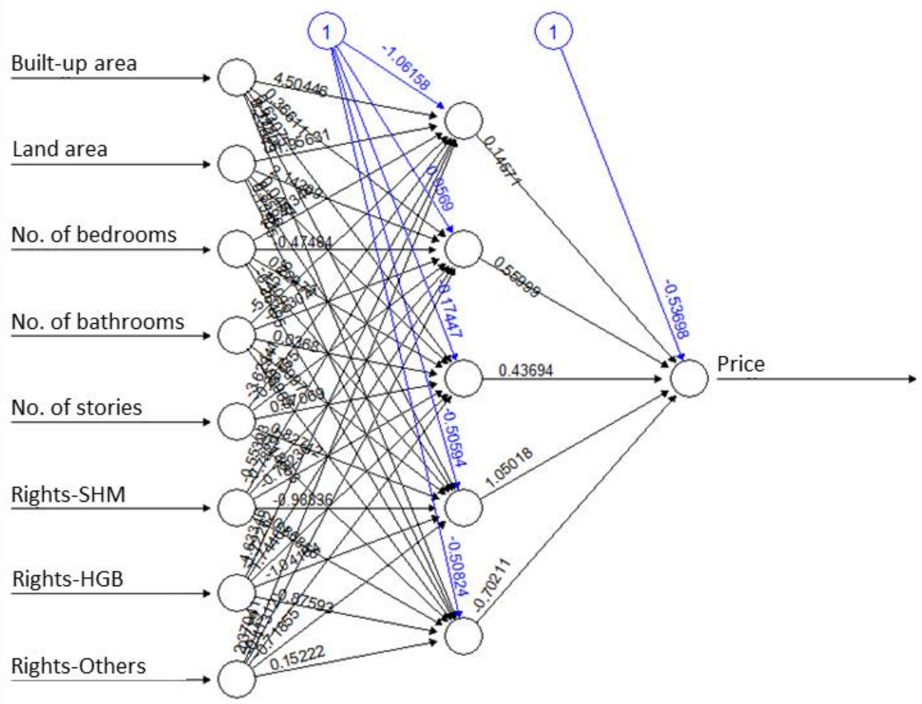


Figure 6. Graphical Representation of ANN House Selling Price

In general, ANN is represented as a collection of layers, namely the input, hidden, and output layers, where each network must have one input layer and one output layer. In this study, the number of neurons in the input layer is the same as the number of predictor variables, and the output layer only has one neuron because the model formed is regression. Meanwhile, in this problem, one hidden layer is considered feasible, with the number of neurons in the layer equal to the average number of neurons in the input and output layers, which is five. After the number of layers and neurons in each layer is determined, ANN will be run on the training data using the backpropagation algorithm.

The graphical representation of the ANN results that have been run can be seen in Figure 6. The black line represents the weight vector between neurons, while the blue line represents the added bias. The backpropagation algorithm training process is always considered a black box in ANN. Hence, the internal characteristics of the network built are a set of weight values that are difficult to explain. However, Garson's algorithm could evaluate the relatives' importance of predictors for house selling prices, as shown in Figure 7. It can be seen that the area of land and buildings, as well as the SHM ownership certificate, are the top three main factors that influence house prices the most.

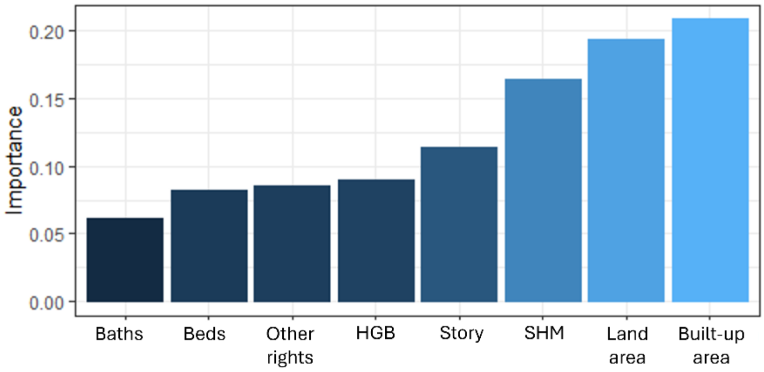


Figure 7. The Relative's Importance of Characteristics of The House

4.4 Performance Comparison of Hedonic Regression Model and ANN

The mass appraisal models of residential property were evaluated using the testing data. A comparison of the prediction results of both methods can be seen visually in Figure 7. The results of a visual inspection of Figure 7 show that the predictions made by ANN are generally more concentrated around the diagonal regression line than the predicted values by the linear model. Thus, the ANN method can predict selling prices closer to the actual price than MRA.

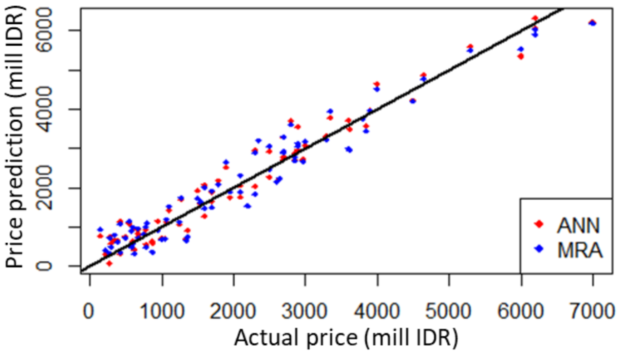


Figure 7. Comparison of Predicted vs. Actual Values from Hedonic Regression and ANN

In addition to visual inspection, model performance evaluation can be measured through error metrics used to calculate the error of the prediction model. The smaller the error metric, the better the prediction model is. Based on Table 4, the ANN has a smaller prediction error on all error metrics than the MRA. It means that ANN generally performs better than MRA.

Table 4. Comparison of Error Metrics Between Hedonic Regression and ANN

Method	Error Metrics		
	MAE	RMSE	MAPE
MRA	$3{,}28 \times 10^2$	$3{,}96 \times 10^2$	34,24%
ANN	$2{,}98 \times 10^2$	$3{,}70 \times 10^2$	31,04%

Although ANN generally performs better than MRA, if we look at the error measurement, ANN still has relatively high errors. Therefore, the ANN model should only be used to support appraisers in assessing property, not as the primary reference.

5. Conclusion

This study applies ANN as an alternative to hedonic regression based on MRA to build a mass appraisal model of residential houses in Surabaya City. The mass appraisal model using hedonic regression has a good coefficient of determination ($adj.\ R\text{-}sq$ = 91%). However, this method cannot capture the nonlinear relationship between house prices and their influencing attributes. Meanwhile, ANN algorithm can be a good alternative in mass property appraisal because of its nature, which can automatically capture nonlinear relationships between input and output variables without determining its nonlinear function. It is backed up by the better performance of ANN than the multiple linear regression method in predicting residential property values.

However, this study has some limitations. First, the selling price of the house used in this study is an estimated price and not the final selling price due to the difficulty of obtaining real sales data from the property market in Surabaya City. Second, the assessment of residential houses in this study has not included external factors such as interest rates, exchange rates, and inflation, as well as environmental factors such as the availability of public facilities that can affect the property value. Therefore, further research must be done before using ANN as a mass appraisal system to assess residential properties in Surabaya City.

6. Acknowledgement

The authors acknowledge and are grateful for the financial support for this work from Institut Teknologi Sepuluh Nopember through a research grant.

References

[1] DJPK, *Modul Penilaian Pajak Bumi dan Bangunan*. Direktorat Jenderal Perimbangan Keuangan, 2018.
[2] MAPPI, *KEPI & SPI Kode Etik Penilai Indonesia dan Standar Penilaian Indonesia*, VII. Jakarta: Masyarakat Profesi Penilai Indonesia (MAPPI), 2018.
[3] V. Kontrimas and A. Verikas, “The mass appraisal of the real estate by computational intelligence,” *Appl. Soft Comput. J.*, 2011, doi: 10.1016/j.asoc.2009.12.003.
[4] N. Anastasia and S. C. Tandono, “Penilaian Apartemen di Surabaya,” *Competence J. Manag. Stud.*, vol. 8, no. 1, 2014.
[5] A. M. Irooth and N. Anastasia, “Model Nilai Pasar Apartemen Dan Kesediaan Membayar View Apartemen di Surabaya,” Petra Christian University, 2017.
[6] L. A. R. Winanda, “Model Penilaian Harga Jual Rumah Kelas Menengah (Tipe 36-45) Dengan Analisis Regresi (Studi Kasus di Wilayah Kabupaten Sidoarjo),” *Prokons J. Tek. Sipil*, vol. 7, no. 2, pp. 155–164, 2013.
[7] K. Adi and M. S. Rifqy, “Analisa Penilaian Harga Jual Rumah Kelas Menengah Dengan Analisis Regresi (Studi Kasus Kabupaten Banyuwangi),” ITN MALANG, 2014.
[8] S. Sucipto, “Penentuan Model Penilaian Harga Jual Produk Perumahan Tipe Menengah Dengan Analisis Statistik Regresi,” *KONTRUKSI*, vol. 1, no. 1, 2013.
[9] S. Yulyanti, “Penilaian Lahan Persawahan Dan Faktor-Faktor Yang Mempengaruhinya Di Kecamatan Ciampel Kabupaten Karawang,” *J. Ekon. KIAT*, vol. 27, no. 2, pp. 95–113, 2016.
[10] N. Firmansyah, “Penentuan Nilai Wajar Aset Tetap Tanah Milik Desa Menggunakan Teknik Penilaian Massal Untuk Tujuan Penyusunan Laporan Kekayaan Milik Desa (Studi pada Desa Dukuhrejo, Kecamatan Bayan, Kabupaten Purworejo),” Universitas Gadjah Mada, 2017.
[11] V. Limsombunchai, C. Gan, M. Lee, and P. O. Box, “House Price Prediction: Hedonic Price Model vs. Artificial Neural Network,” *Am. J. Appl. Sci.*, vol. 1, no. 3, pp. 193–201,

2004.
[12] B. Božić, D. Milićević, M. Pejić, and S. Marošan, “The use of multiple linear regression in property valuation,” *Geonauka*, vol. 01, no. 01, pp. 41–45, 2013, doi: 10.14438/gn.2013.06.
[13] P.-Y. Lai, “Analysis of the Mass Appraisal Model by Using Artificial Neural Network in Kaohsiung City,” *J. Mod. Account. Audit.*, vol. 7, no. 10, pp. 1080–1089, 2011.
[14] C. C. Lin and S. B. Mohan, “Effectiveness comparison of the residential property mass appraisal methodologies in the USA,” *Int. J. Hous. Mark. Anal.*, 2011.
[15] F. Fahirah, “Identifikasi Faktor Yang Mempengaruhi Nilai Jual Lahan Dan Bangunan Pada Perumahan Tipe Sederhana,” *SMARTek*, vol. 8, no. 4, 2010.
[16] H. Supriyono, “Pemilihan rumah tinggal menggunakan metode weighted product,” *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 1, no. 1, pp. 23–28, 2015.
[17] M. A. Shahin, M. B. Jaksa, and H. R. Maier, “Recent Advances and Future Challenges for Artificial Neural Systems in Geotechnical Engineering Applications,” *Adv. Artif. Neural Syst.*, 2009, doi: 10.1155/2009/308239.
[18] J. Zurada, A. S. Levitan, and J. Guan, “A Comparison of Regression and Artificial Intelligence Methods in a Mass Appraisal Context,” *J. Real Estate Res.*, vol. 33, no. 3, pp. 349–388, 2011.
[19] C. Dewi, “Perbandingan Nilai Tanah Menggunakan Model Analisis Regresi Berganda dan Jaringan Syaraf Tiruan (Studi kasus: Kelurahan Way Lunik, Ketapang dan Way Laga, Kota Bandar Lampung),” *Rekayasa J. Ilm. Fak. Tek. Univ. Lampung*, vol. 15, no. 3, pp. 175–182, 2011.
[20] A.-A. Mousa, “An evolutionary-optimised artificial neural network approach for automatic appraisal of Jordanian lands and real properties,” in *Proceedings of the 11th WSEAS international conference on nural networks and 11th WSEAS international conference on evolutionary computing and 11th WSEAS international conference on Fuzzy systems*, 2010, pp. 203–208.