Vol. 4 No. 1 (2025)
e-ISSN: 2963-7589
Economic and Business Journal | ECBIS
https://ecbis.net/index.php/go/index

DATA-DRIVEN MARKETING IN REAL ESTATE: FORECASTING HOUSE PRICES AND UNCOVERING INFLUENTIAL FACTORS

Intania Prabadianti[1], Samidi[2]
[1-2)] Master of Science in Management Program, Faculty of Economics and Business, Padjadjaran University
E-mail: [1)]intania20001@mail.unpad.ac.id, [2)]samidi@budiluhur.ac.id

Submitted:	Revised:	Accepted:
15 September 2025	19 October 2025	21 November 2025

Abstract

In recent years, the housing market has faced significant challenges, including fluctuating prices and declining sales. To solve this issue, there was an increasing need for more sophisticated methods to predict housing prices accurately. This study aimed to provide real estate marketers with a tool to enhance their pricing tactics and mitigate the decline in home sales by predicting house prices using machine learning techniques. Several parameters were considered in this study, such as location, number of bedrooms, number of bathrooms, land area, building area, and number of carports. Linear regression and neural network methods were used to develop predictive models. The findings showed that the neural network method was more accurate than linear regression, which made it a better tool for real estate pricing strategies, with land area and number of carports being the most influential aspects in house price prediction.

Keywords: *house price prediction, machine learning, data-driven marketing, linear regression, neural net.*

INTRODUCTION

Homes are among the most fundamental necessities of human existence, besides food, water, and even the foundation of family stability and well-being (Q. Wang et al., 2024; Zulkifley et al., 2020). It pertains to the concept of homes as property that most individuals in society incur costs for in terms of ownership and upkeep (Jafari & Akhavian, 2019). The housing sector is a significant driver of economic growth in developing and established nations since housing is the largest asset and expense (F. Wang & Ran, 2019). The quality of life for all households is significantly influenced by some aspects of their home, which can directly impact their overall well-being (Bottini & Costarelli, 2023). Everyone wants to buy and live in a house that matches their lifestyle and provides amenities that meet their desires; however, predicting the price of a house is challenging since it is constantly changing (Sanyal et al., 2022). Thus, choosing a price for purchasing and selling a house is a procedure that must be thoroughly examined (Jáuregui-Velarde et al., 2023).

House price forecasts are crucial for many individuals and organizations (H. Zhang et al., 2023) since house prices have become a significant issue over time (Whelan et al., 2023). Knowing the projected house price could help house buyers plan better for real estate purchases, and it is good for house sellers to be aware of the estimated house price on the market to propose a fair asking price (H. Zhang et al., 2023). In the real estate field, it is beneficial for potential buyers of homes to assess the value of houses before making any decision, and it acts as a reference for developers to establish appropriate housing prices, considering various criteria such as location, size, and accessibility of the property (Teoh et al., 2023). Precisely forecasting the price of a house can assist purchasers and other decision-makers in making informed choices (Jáuregui-Velarde et al., 2023). Inaccurate house price predictions increase the level of uncertainty and risk for both homeowners and investors, potentially resulting in financial losses due to unforeseen market fluctuations (Gandhi et al., 2023; Geerts et al., 2023; Mishra et al., 2022). The effectiveness of the real estate market largely depends on reliable price forecasting. When such forecasts are lacking, the market may suffer from inefficiencies, mispricing, and poor investment outcomes (Geerts et al., 2023; Verma et al., 2023).

In recent years, however, the housing market has faced significant challenges, including fluctuating prices, economic downturns, and changes in consumer behaviour (Y. Zhang et al., 2023; Zulkifley et al., 2020) One of the most pressing issues is the decline in home sales observed in various regions, and unfortunately, Indonesia is facing this issue. The volatility of housing prices has consistently been a prominent subject of study in building, real estate, and built environment research (Jafari & Akhavian, 2019). According to Indonesia's residential property price index year on year, the price has fluctuated from the 1st quarter of 2022 to the 4th quarter of 2023. It can be increased by 1-2% but decreased by 1-2% over the observed period (Statista, 2024). Similar to the property price index growth, Indonesia's year-on-year residential property sales growth from the 1st quarter of 2020 to the 2nd quarter of 2023 also experienced fluctuation. From the 2nd quarter of 2022 until the 2nd quarter of 2023, sales kept going down, and overall property sales decreased by 12.3%. Among the three housing types, the sales for small houses are the most significant overall decline (Statista, 2023).

The decrease in home sales can be attributed to various factors. Traditional market analysis techniques and pricing methodologies sometimes need help to analyze large amounts of data, resulting in underutilization (Q. Wang et al., 2024). This insufficiency in pricing strategy causes the sales to decline, producing a tough-to-break cycle. To address this issue, there is a growing need for more sophisticated methods to predict housing prices accurately. Accurate price prediction can help businesses utilize the forecasted house price value to modify their strategy to assist those experiencing economic obstacles in owning a home (Y. Zhang et al., 2023). Housing data sets are typically complex, containing numerous feature attributes. Therefore, there is a lack of clarity regarding which attributes are the most significant, and machine learning (ML) excels in selecting the appropriate factors to consider, which is often challenging (Teoh et al., 2023). Machine learning techniques offer promising solutions in this regard.

The key objective of this study is to develop and implement ML models for predicting housing prices. The model will utilize various features such as location, number of bedrooms, bathrooms, carports, land area, and building area to generate accurate price predictions. By doing so, the study aims to provide real estate professionals with a tool to enhance their pricing strategies and mitigate the decline in home sales. This research holds significant implications for the real estate industry. Firstly, it can help real estate marketers set more accurate and competitive prices, attracting more potential buyers. Secondly, stakeholders can better address market demands and preferences by understanding the key influencing factors in housing prices. Finally, implementing ML in this context could pave the way for broader ML applications in real estate, promoting the advancement of novel ideas and improved efficiency within the sector. This study also differs from past studies in that it focuses on analyzing data specifically from properties for sale in Bandung, which adds a fresh aspect to the study. In contrast to earlier studies, this research primarily focuses on the main components of a house and does not consider environmental variables. Research also supported by a literature review that outlines relevant theories and previous research as the basis for the analysis, namely as follows:

Machine Learning for House Price Prediction

Artificial intelligence (AI) generally refers to robots or agents that can observe their surroundings, acquire knowledge, and make intelligent judgments or suggestions based on that knowledge; one advanced AI is Machine Learning, which can be used for recognizing, understanding, and analyzing data (Rampini & Re Cecconi, 2021). Within machine learning, there are clear-cut classifications of learning, namely supervised and unsupervised learning (Kaytan and Aydilek 2017). Supervised learning contains several models such as linear regression (LR), neural network, support vector regression (SVR), ensemble approach, decision tree (DT), Naive Bayes, K-nearest neighbour (k-NN), and discriminant (Hoxha, 2024).

Machine learning algorithms have become popular for house price prediction (Y. Zhang et al., 2023) since they can identify the crucial variables contributing to house prices (Teoh et al., 2023). ML algorithms utilize sample data to predict housing prices by building a model through experience rather than using explicit modeling equations that show how dependent and independent factors are connected. Several supervised learning techniques have been utilized in prior studies to forecast housing prices (Adetunji et al., 2021; Forys, 2022; Y. Zhang et al., 2023). The goal of supervised learning is to establish a model that accurately represents the connection between input and output variables (Kotu & Deshpande, 2019).

Normal regression problems are typically addressed with the linear regression model as the baseline model (Trang et al., 2021). Zhao et al. (2019). Developed a highly effective price prediction model using linear regression, random forest, neural network, and XGBoost methods. Other research that uses linear regression, random forest, and decision tree methods also generates good accuracy, with accessibility as the most influential housing price factor among physical condition, socioeconomic characteristics, built-form, and location (Y. Zhang et al., 2023). (Adetunji et al., 2021) also applied the random forest method to help people decide whether the property should be acquired by considering the' three most important factors of house prices: physical condition, style, and location.

Zulkifley et al. (2020) analyzed by building upon prior studies and derived a valuable summary of the key attributes that hold the most significance in housing analysis, which revealed that support vector regression, artificial neural networks (ANN), and XGBoost were the three most effective methods for predicting house prices with location and structural attributes as the prominent factors in predicting house prices and the locational characteristic was shown to be the primary element influencing the accuracy of the predictions, resulting in low error rates. Research conducted by (Rampini and Re Cecconi, 2021) shows that in predicting house prices, the surface and box of the parking lot are the most important factors, and between the use of ElasticNet, XGBoost, and neural networks that were tested, the one that performed better than the others is neural networks. Some of the previous research also proves that Neural networks are the best fit for predicting house prices (Forys, 2022). Based on existing research, this study proposes linear regression as the basic method always used in price prediction and the neural networks method, which has recently become reliable.

Factors in House Pricing

Several studies have analyzed the factors that influence home affordability from both micro and macro viewpoints (Q. Wang et al., 2024). Previous research, which tested through correlation and regression analysis, considers the macroeconomic indicators as a strong link between house price stability by the inflation rate, GDP, income level, and employment level that influence the affordability of house prices (Sabrina Abdul Latif et al., 2020). According to

Zheng (2023), GDP growth combines with education, health insurance, household, and income. Most of the other research involves more of the micro point, which are the factors that match the preferences of potential buyers. These factors include elements that encompass the attributes of the house, such as the livable surface, number of bedrooms and bathrooms; location area that refers to specific facts related to the geographical position, such as closeness to parks, supermarkets, and universities; and socioeconomic aspects, such as the average income and education level of its residents (Lenaers et al., 2024).

House prices are primarily influenced by accessibility to job centers and facilities, as well as the environmental qualities of the region (Kang, 2025; Szumilo et al., 2017; van Ruijven & Tijm, 2024). According to research conducted by Forys (2022), property appraisers stated that the basic characteristics of the land property are of priority, such as the property's condition, neighborhood, plot area, distance from the city center, and location. Other research categorized the factors that influence housing prices as locational, structural, and neighbourhood factors (Zulkifley et al., 2020). Similarly, (Wu et al., 2018) also determined that the influencing factors of housing price consist of location features, neighborhood, and architecture, which were later added by land-use planning, environment, block size, and wind power. Despite both of the factors seeming to have a lot of influence on housing prices, the macroeconomics factor is less frequently employed (Zulkifley et al., 2020). The rationale behind this finding is that including macroeconomic variables in prediction models can be more challenging due to their weaker correlation with real estate prices than location factors and property attributes (Warisse, 2017). Another reason for this is the unclear and contradictory research findings on the correlation between the analyzed macroeconomic factors and property prices (Chiwuzie, Dabara, and Omotehinshe, 2021). Therefore, the structural and location factors are the two principal categories (Krämer et al., 2023). Following the microeconomic approach, this study concentrates especially on the structural and locational features of the house, such as location, land area, building area, number of bedrooms, number of bathrooms, and carports, which are regarded as more directly measurable and influential in property pricing.

2. RESEARCH METHOD

The Cross Industry Standard Process for Data Mining (CRISP-DM), which was created by a consortium of organisations specialising in data mining, is a prevalent framework for data science processes widely used to produce solutions in the field (Kotu and Deshpande, 2019). CRISP-DM organises the data mining process into six phases: business understanding (project goals and requirements), data understanding (data collection and insight discovery), data preparation (data selection and transformation, modeling (use of techniques), evaluation (assessing the model), and deployment (monitoring and presenting gained knowledge). These phases assist businesses in comprehending the data mining process and offer a roadmap for planning and executing a data mining project. Here is the research flow

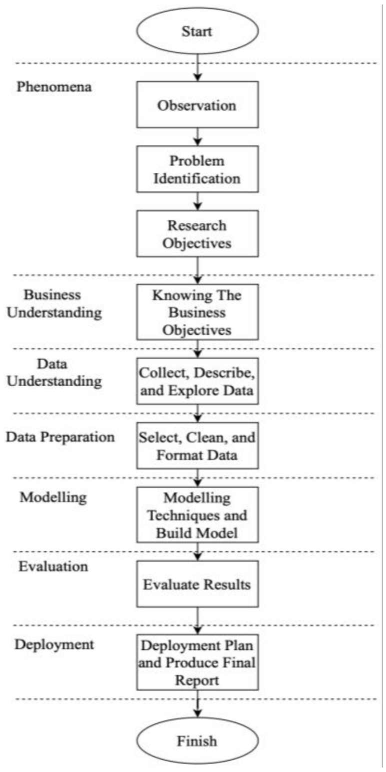

Figure 1. Research Flow

The dataset comprises diverse numerical data that will undergo quantitative analysis, including linear regression and Neural networks, to forecast housing prices and identify the primary factors influencing them. The study will be conducted using the RapidMiner Studio 10.3 software to run linear regression, ANN, and performance evaluation methods.

2.1. Linear Regression

Linear regression is a technique used for numeric prediction in which the main concept is to forecast the value or class of a dependent characteristic, y, by creating a function, y=f(X), that combines the predictor attributes, X (Kotu and Deshpande, 2019). Hence, regression analysis is a model that determines the relationship between variables (Zulkifley et al., 2020). The regression approach is widely used to model the house price affordability index drivers. When conducting research using regression methods, it is typical to encounter challenges such as multicollinearity, which arises from the interdependence of variables, as well as dealing with a large number of variables (Q. Wang et al., 2024). The linear regression model takes the form:

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (1)$$

Linear regression is still utilized as a house price prediction model despite being the most basic model compared to others, which, with this method, allows for precise training to get the desired outcomes (Trang et al., 2021). This is because the linear regression machine learning approach applies a linear regression model to make predictions within a continuous range rather than discrete categories (Kavitha et al., 2016). The linear regression model serves as a baseline for evaluating the effectiveness of various machine learning models by comparing their performance metrics, such as MAE, MSE, and R-square scores, to identify the ideal model (Trang et al., 2021).

2.2. Artificial Neural Networks

Artificial Neural Networks (ANNs) consist of a sequence of trainable modules called layers, where these layers do not need any parameter adjustment except for the network size (Lecun et al., 2015). ANN is designed to mimic the functioning of biological neural systems, such as the brain's information processing. They consist of many interconnected processing units called neurons, which collaborate to solve specific issues (J. Zhao & Zhang, 2021). The ANN is consistently chosen when dealing with non-linear attributes (Zulkifley et al., 2020). This method is becoming increasingly popular due to its high efficiency and straightforward implementation; also, the most notable advantage of neural networks is that they require little human intervention in tweaking the model; thus, the neural network is relatively easy to apply (Rampini & Re Cecconi, 2021). The inner workings of the neural network modeling process are explained by the developers of this technique using various biological terms; however, it is based on a straightforward mathematical framework that is:

$$Y = 1 + 2X_1 + 3X_2 + 4X_3 \quad (2)$$

Y is the estimated outcome, and X1, X2, and X3 are the input characteristics. The intercept is 1, whereas the scaling factors or coefficients for the input characteristics X1, X2, and X3 are 2, 3, and 4, respectively (Kotu & Deshpande, 2019). ANN has three layers consisting of the first layer, in which data is introduced into the model; the second or middle layer, which is the “hidden layers” that determine the correlations between features to achieve the most accurate forecast; and the last layer is where the results are provided in the form of a prediction value for the target variable (Rampini & Re Cecconi, 2021). The output results of the ANN model display the relative weights through colour-coded links with the description tab in the model window, providing the precise values of the link weights (Kotu & Deshpande, 2019).

2.3. Performance Evaluation

Forecast accuracy standards must be developed to assess the performance of different models and compare them across various use cases (Kotu & Deshpande, 2019). With the advancement of computer technologies, researchers are increasingly faced with the challenge of determining the accuracy of a specific prediction algorithm (Edalatpanah et al., 2024). The utilization of R-squared (R2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) as assessment metrics highlights the significance of employing comprehensive and diverse criteria to assess the performance of a model (Hoxha, 2024). MAE is a straightforward measurement influenced by the data's scale (Chai, 2020). MAE is commonly favoured due to its ability to represent the average error accurately, and a higher number of MAE indicates a decrease in the method's accuracy. The maximum accuracy, which aligns with the actual value, is achieved when the MAE is zero (Rampini & Re Cecconi, 2021). MAE is defined as follows:

$$MAE = \frac{\sum_{i=1}^{m} (Yi - \hat{Y}i)}{m} \quad (3)$$

Root Mean Square Error (RMSE) is a measure that depends on the scale of the data and is employed when it is important to penalise high relative residue (Kotu & Deshpande, 2019) RMSE is employed to assess the accuracy of forecasts where the variability of the errors in predicting values (residuals) is calculated (Sanyal et al., 2022). RMSE is defined according to the following formulas:

$$RMSE = \frac{1}{m} \sum_{i=1}^{m} \left(Yi - \hat{Y}i\right)^2 \quad (4)$$

Where m represents the total number of observations, including the train and test sets used to evaluate models. The actual rent for the property is denoted as Yi, whereas Ŷi represents the forecast rent for the property (Lenaers et al., 2024). R2 is also considered a good measure for the fitness evaluation with ranges from 0 to 1, where the closer it is to 1, the more accurate it is (Sanyal et al., 2022) The formula that is used for R-squared is:

$$R^2 = 1 - \frac{SSE}{SST} \quad (5)$$

SSE (Squared sum of error) is the sum of squared residuals calculated from the difference between each observation and its corresponding anticipating value, and SST (Sum of squared total) is the squared differences between each observation and the aggregate mean (Sanyal et al., 2022).

3. RESULTS AND DISCUSSION

3.1 Business understanding

Real estate marketers and property developers are challenged when house sales decline in a competitive property market. To enhance the efficacy of sales and marketing activities, a primary business goal was to create a precise predictive model for evaluating house prices using different property attributes and determining the aspects that have a substantial impact. Using this strategy, the company aims to establish more favourable pricing, gain insights into market preferences, and enhance sales volume.

3.2 Data Understanding

The initial step in the data comprehension phase involves gathering relevant information for analysis. The data for this research were obtained through web scraping of www.rumah123.com in March 2024, specifically focusing on residential properties in Bandung, West Java, Indonesia. The raw datasets consist of type, status, price, instalment, house name, location, bedroom count, bathroom count, carport count, land area, and building area. Each of the raw data categories is described as

Table 1. Description of Raw Dataset

Data	Data Type	Description
Type	Nominal	The type or classification of the property
Status	Nominal	The property’s status (available or not)
Price	Numeric	The property’s price in Indonesia Rupiah (IDR)
Instalment	Nominal	Any instalment plan information related with the property
House name	Nominal	The name or title of your residential property
Location	Nominal	The location or area of the propery in Bandung
Bedroom count	Numeric	Number of bedrooms in the property
Bathroom count	Numeric	Number of bathrooms in the property
Carport count	Numeric	The amounts of carports or parking places available on the site
Land area	Numeric	The property’s total land area measured in square meters
Building area	Numeric	The property’s total building area measured in square meters

Source: Data Processed, 2025

These datasets are intended to facilitate the study and analysis of real estate, specifically housing price regression, in Bandung, West Java, Indonesia.

3.3 Data Preparation

The initial step in the data preparation phase involves selecting pertinent data for analysis. The information utilized for the house price analysis comprises variables often used in previous research: location, bedroom count, bathroom count, carport count, land area, building area, and price (Dwivedi et al., 2022; Lenaers et al., 2024; Njo et al., 2025). The data was chosen due to its potential to substantially impact housing prices and its ability to offer comprehensive insights for predicting prices. After selecting the data, the next task is to cleanse it to ensure its quality and consistency. The data cleansing procedure entails verifying and addressing missing values, correcting data inaccuracies (such as typographical errors or illogical values), and eliminating duplicate entries.

Once the data has been cleaned, the next step is to format it to prepare it for analysis. This includes data type conversion, standardisation of numerical values to ensure a consistent scale, and categorical data encoding. Proper data formatting ensures that analysis algorithms can work optimally. The current data is guaranteed to have data categories that align with the aim of the data itself and is standardised using the "normalize" operator. During this phase, the location data is also transformed from a nominal to a numerical format using the "nominal and numerical" operator. This is achieved by specifying the attribute filter type as a subset and the coding type as unique integers. As a result, the original location data, which consisted of sub-district names in Bandung, is converted into a numerical format ranging from 0 to 26. Each number corresponds to a different sub-district. The clean dataset has 6534 data points and eight columns. The columns include row number, price, location, bedroom count, bathroom count, carport count, land area, and building area. The price data serves as a label or dependent variable. The result of the executed operation produces the following data set. To provide an overview of the dataset structure and its statistical characteristics, the following tables present the data profile, including data types, minimum and maximum values, mean, and standard deviation, as well as the first ten rows of the cleaned dataset.

Table 2. Data Profile

Data	Data Type	Minimal	Maximal	Mean	Std. Deviation
Price (in million)	Numeric	170	22000	3206.876	3364.787
Location	Nominal	-	-	-	-
Bedroom count	Numeric	1	15	3.809	1.806
Bathroom count	Numeric	1	10	2.652	1.300
Carport count	Numeric	0	10	1.206	1.221

Land area	Numeric	36	687	201.993	138.993
Building area	Numeric	36	1126	205.476	150.313

Table 3. First 10 Rows of Clean Dataset

Row No.	Price (in million)	Location	Bedroom count	Bathroom count	Carport count	Land area	Building area
1	2100	0	3	2	2	137	170
2	4100	0	3	2	3	202	300
3	3300	0	5	2	1	350	258
4	1300	0	11	3	0	176	176
5	3600	0	5	3	1	184	234
6	1170	0	2	1	1	65	45
7	2900	0	4	3	2	260	260
8	4500	0	5	3	2	404	250
9	2100	0	3	2	2	137	170
10	2200	0	4	3	2	172	220

3.4 Modelling

3.4.1. Correlation Matrix

The correlation matrix generated in RapidMiner for the housing dataset offers valuable insights into the associations among several features, such as the number of bedrooms, bathrooms, carports, land area, building area, and the goal variable, house price. By analysing correlation coefficients, we may determine the features that exhibit the most robust linear associations with property prices. A correlation value near +1 suggests a strong positive linear relationship. The correlation matrix data above shows that the land_area variable has a very strong relationship with building_area, with a correlation value of 0.749. This shows that houses with larger land areas tend to have larger building sizes as well, both of which can be important indicators in determining house prices. In addition, bathroom_count also shows a fairly strong correlation with building_area (0.590), followed by bedroom_count with building_area (0.496), indicating that larger homes are generally equipped with more bedrooms and bathrooms. A moderate relationship is also seen between bathroom_count and land_area (0.446). On the other hand, the location variable shows a very weak correlation with the other features, with values below 0.1, so its contribution is expected to be insignificant in the context of house price prediction, its contribution is expected to be insignificant. Based on these results, variables land_area, building_area, bedroom_count, and bathroom_count have great potential to be used as the main predictors in building a house price estimation model.

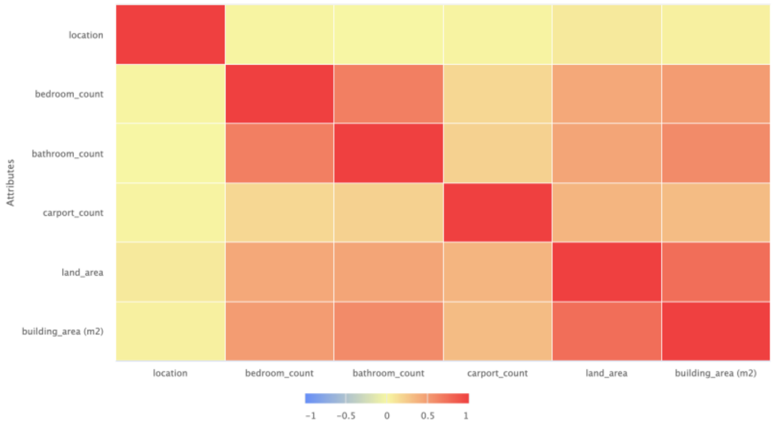

Figure 2. Correlation Matrix

3.4.2. Linear Regression

The Linear Regression method was employed to forecast house prices using the prepared dataset, which has columns labelled location, number of bedrooms, number of bathrooms, number of carports, land area, building area, and price. The results are as follows.

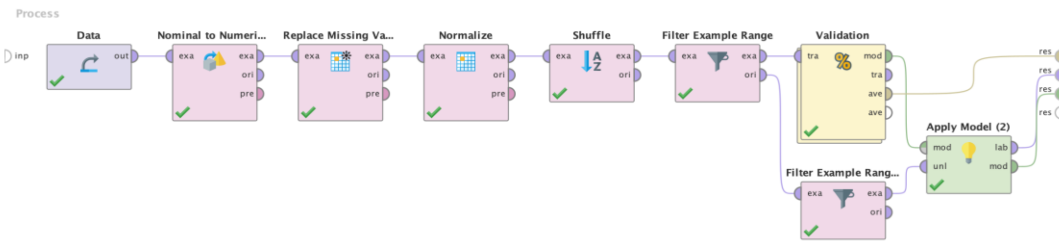

Figure 3. First Process of Linear Regression

The "normalize" and "replace missing values" operators will be connected to the clean dataset to guarantee that the data is standard and there are no null values. Subsequently, employ the Shuffle operator to randomize the order of the data, ensuring that the two partitions are statistically equivalent when separated.

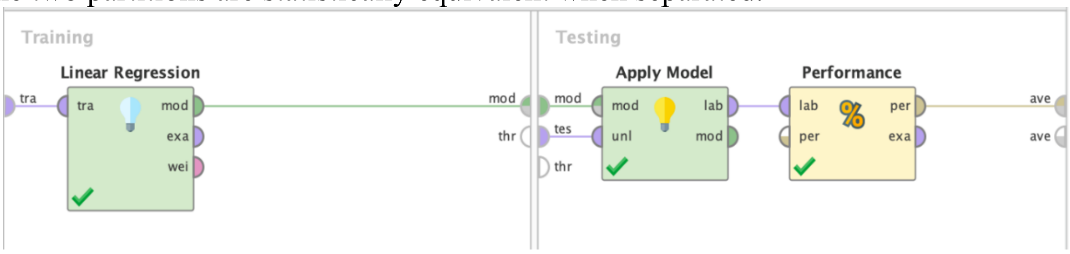

Figure 4. Second Process of Linear Regression

Next, divide the data into two sets using the Filter Examples Range operator. In this instance, 90% of the data is designated for training purposes, while 10% is designated for testing purposes. The training data will be further divided into training and validation sets. The default Split Validation options, which include relative, 0.7, and shuffled, will remain unchanged (Kotu & Deshpande, 2019), and this will be required to evaluate the linear regression model's performance. In addition, it is advisable to establish the local random seed, which defaults to 1992. This guarantees that RapidMiner will select the same samples if the procedure is executed later. The linear regression operator is included in the validation and is used to apply the model. The following is a graph of the price prediction results compared to existing prices.

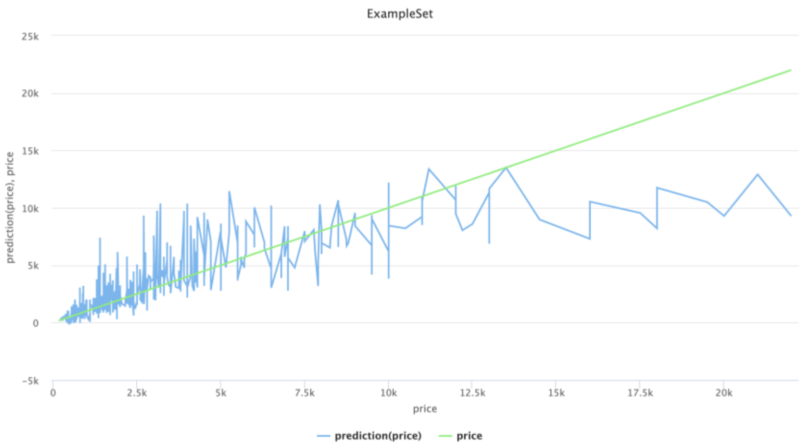

Figure 5. Comparison of Price Prediction and Existing Price in Linear Regression

According to the linear regression results, the land area, building area, and number of carports have a significant influence on house prices. This is indicated by the p-value of the three variables, which are below the 5% significance level ($p<0.05$), each amounting to 0.000. The land area variable shows a regression coefficient of 1,739.790, which shows that each increase in one unit of land area has the potential to increase house prices by Rp1,739,790. Meanwhile, the building area has a coefficient of 1,176.987, and the coefficient of the number of carports is 196.486, which means that an increase in building area and carports is also positively correlated with an increase in house prices. The results of linear regression indicate that land area, building area, and number of carports strongly influence the prices of houses. This result is in line with research conducted by Hamami & Dahlan (2024). In contrast, location and number of bedrooms weakly influence the house price prediction. The equation resulting from this result is

$$House\ Price = (22.725 * location) + (-43.756 * bedroom\_count) + (-62.584 * bathroom\_count) + (196.486 * carport\_count) + (1739.790 * land\_area) + (1176.987 * building\_area) + 3217.981 \quad (6)$$

3.4.3. Artificial Neural Networks

Another technique utilized for predicting property prices is neural networks. The operator components used are identical to those employed in linear regression up until the point of data normalization. Following the application of the "normalize" operator, the subsequent step involves using the "split data" operator to partition the data into training and testing data, with a ratio of 90% for training and 10% for testing. Lastly, the performance regression operator is also implemented. Below is the proposed method design, and the hidden layer consists of five nodes that use the sigmoid activation function. The result of this hidden layer is the weight of each input characteristic on a node, which represents its level of influence.

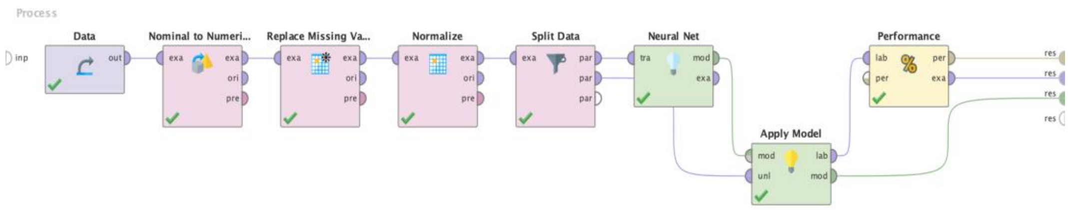

Figure 6. Neural Net Process

Table 4. Hidden Layer Results

Factors	Node 1	Node 2	Node 3	Node 4	Node 5
Location	-0.004	0.362	-0.021	-1.313	1.278
Bedroom count	-0.246	2.535	0.404	0.222	1.305
Bathroom count	0.334	0.715	-1.197	0.763	0.695
Carport count	0.079	4.538	-0.665	1.623	2.309
Land area	-1.431	0.337	-1.041	0.328	-0.452
Building area	-1.845	2.415	0.215	0.854	-0.284
Bias	-0.624	-0.141	-3.209	-1.988	-2.239

Node 1 demonstrates that the size of the building and the land area have a strong negative impact on housing prices, while the number of bathrooms and carports exerts a positive influence. The number of bedrooms and location show minimal effect in this node. Node 2 demonstrates that the number of carports, building area, and bedrooms have a strong positive impact on pricing, followed by location and bathroom count. In contrast, the influence of land area is positive but relatively moderate. Node 3 highlights a significant negative impact from the bathroom count, carport count, and land area, while the building area contributes slightly positively. The influence of location and bedrooms is present but weaker. Node 4 indicates that the number of bathrooms, bedrooms, carports, land area, and building area contribute positively to house prices, while the location exerts a strong negative impact. Node 5 demonstrates that the location,

ECBIS ECONOMICS AND BUSINESS JOURNAL
Vol. 4 No. 1 (2025)
e-ISSN: 2963-7589
Economic and Business Journal | ECBIS
https://ecbis.net/index.php/go/index

number of bedrooms, bathrooms, and carports have a strong positive impact, whereas land area and building area have a mild negative influence.

In general, Node 2 exhibits a substantial positive effect on the number of carports and building area, while Node 3 shows negative effects from bathroom count and land area. Overall, features such as the number of bedrooms, bathrooms, carports, building area, and land area substantially influence house values, whereas location's impact varies across nodes. These findings are consistent with previous research conducted by (Jafari & Akhavian, 2019). While the neural network model in this study reveals inconsistent patterns in the influence of location across different hidden nodes, previous studies have shown that geographic location, particularly at the district or subdistrict level, can significantly affect house prices, especially in more developed areas (Mukhlishin et al., 2017; Njo et al., 2025; Rahmatillah & Sudirman, 2023). The image below illustrates the comparison between the expected and existing prices, as seen in the chart.

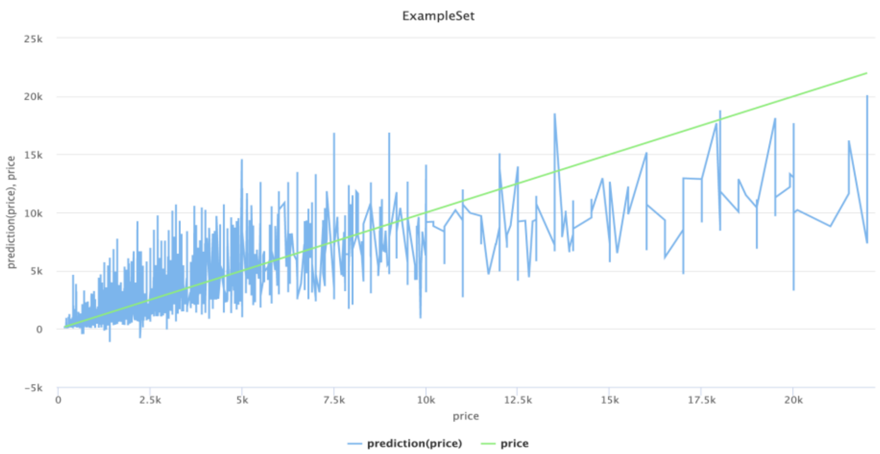

Figure 7. Comparison of Price Prediction and Existing Price in Neural Net

3.5 Evaluation

The selected performance criteria are RMSE, MSE, MAE, and R-squared. Performance is then evaluated using these metrics. Here are the results of the linear regression method and neural net method.

Table 5. Performance Evaluation

Method	RMSE	MAE	R-squared
Linear Regression	1931.063	1151.113	0.673
Neural Net	1894.525	1059.474	0.701

The fairly high RMSE shows that the model captures the overall trend well, but there are big differences between some of its predictions. The neural network model achieved better predictive performance, as indicated by a lower RMSE of 1894.525 compared to 1931.063 for linear regression, suggesting that its predictions are generally closer to the actual house prices. Similarly, the neural network also recorded a lower MAE of 1059.474 versus 1151.113, indicating smaller average absolute errors. Regarding R-squared, the neural network explains 70% of the variance in house prices, slightly outperforming the linear regression model with a score of 0.673. These results suggest that the neural network is more effective in capturing complex patterns and non-linear relationships in the data.

3.6 Deployment

During the Deployment phase, the initial step involves strategizing and integrating the house price prediction model into a functional environment. This entails incorporating the model into technologies used by end users, such as property sales platforms or real estate management systems. For practical application, the predictive model can be embedded within a web-based interface, allowing real estate agents or prospective buyers to obtain real-time house price estimates by entering relevant property details. This feature can be beneficial not only for buyers in assessing whether a house is reasonably priced based on its specifications, but also for sellers or real estate agents in setting an appropriate asking price when putting a property on the market. Beyond implementation, it is essential to establish a structured approach for ongoing monitoring and maintenance. This includes consistently evaluating model performance using key metrics such as RMSE, MAE, and R-squared to ensure predictive accuracy and reliability. Additionally, a protocol must be in place for handling periodic data updates, such as integrating new property listings, so the model can be retrained when necessary to reflect evolving market conditions. To ensure sustained relevance and accuracy, the deployment process should also account for potential issues such as data drift or changes in user behavior. Establishing a feedback loop from end users can further enhance the model by aligning its outputs with practical needs and expectations in the real estate sector.

4. CONCLUSION

The results of this study provide practical insights for real estate marketers to improve pricing accuracy and campaign effectiveness. Predictive models such as neural networks allow marketers to reduce the risk of overpricing or underpricing properties, which is essential for faster sales, as well-priced homes tend to attract more prospective buyers (Dwivedi et al., 2022; Saraswathi et al., 2024). Based on the neural network results, three structural features, which are carport count, building area, and number of bedrooms, were identified as the most influential factors in determining house prices. These findings align with previous studies emphasizing the importance of physical attributes such as size, number of rooms, and overall condition of the property in influencing market value (Hamami & Dahlan, 2024). Marketers can leverage this insight by highlighting these high-impact features in their listings and campaigns, for example, promoting homes with ample living space, multiple carports, or additional bedrooms tailored to family-oriented buyers.

Although land area and bathroom count also play roles in pricing, their impact in the neural network model was less consistent across hidden layers, suggesting that these features may not universally raise perceived value. Therefore, marketers are advised to avoid one-size-fits-all assumptions and instead tailor promotional strategies based on the most relevant features for specific property segments. Another key implication lies in the integration of predictive pricing tools into real estate platforms, which allows marketing teams to automate price recommendations for new listings using reliable data, which can reduce subjectivity and improve consistency in pricing decisions, resulting in more targeted and data-driven marketing strategies (Y. Zhao et al., 2019). This could represent an upgrade in terms of customer experience, which would positively influence behaviour intention (Eny et al., 2020). As a result, aligning marketing efforts with data-driven predictions strengthens the property's market position and enhances both customer satisfaction and sales performance.

The study successfully utilised the Linear Regression and Neural Net models to forecast house prices using characteristics such as location, number of bedrooms, number of bathrooms, carport, land area, and building area. The variables equally identified as positive and significant determinants in both methods are the land area and the number of carports. Previous research states that general living area, basement size, and remodelling age are the most significant factors in determining house prices in America. In contrast, in Australia, properties with bedrooms, bathrooms, land area, and distance to the CBD are the main factors in house prices (Teoh et al., 2023). The number of carports that influence house prices is also supported by the research conducted by Rampini & Re Cecconi (2021). According to the performance findings, both proposed models show an R-squared value of 67,3% for linear regression and 70,1% for the neural network. Aligning with this result, Artificial Neural Networks (ANNs) demonstrate superior performance in house price prediction compared to other models, as they can identify complex and non-linear patterns within the data (Njo et al., 2025). Additionally, the error factor in the neural network model is lower than the error factor in the linear regression model. These findings demonstrate that the neural network approach is preferable to linear regression for predicting house prices, especially with this dataset.

This study has several limitations. Firstly, the data used may not account for all variables that influence property values, such as macroeconomic conditions, income, and the quality of the neighbourhood, including the availability of various amenities. Second, the linear regression model employed in this study might not adequately capture the complex non-linear relationships that often exist between the variables. Real estate markets are influenced by multifaceted interactions among various factors, and a linear approach may need to be more accurate to justify these dynamics. Furthermore, using region-specific data, such as the names of sub-districts, introduces another layer of complexity, making it more difficult to interpret the results universally. To enhance future research, it is advisable to include more complex non-linear models, such as Random Forest Regression, which can effectively capture non-linear interactions. Furthermore, enhancing the dataset by incorporating additional variables, such as the availability of public facilities, environmental circumstances, and economic trends, could improve the precision of predictions. Ultimately, establishing an ongoing monitoring and maintenance system will guarantee the model's continued relevance and accuracy in response to fluctuations in the property market.

REFERENCES

Adetunji, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., & Oluwadara, G. (2021). House Price Prediction using Random Forest Machine Learning Technique. *Procedia Computer Science*, *199*, 806–813. https://doi.org/10.1016/j.procs.2022.01.100

Bottini, L., & Costarelli, I. (2023). Housing Satisfaction between Individual and Regional Characteristics: The Italian Case; [La soddisfazione abitativa tra caratteristiche individuali e regionali: il caso italiano]. *Scienze Regionali*, *22*(2), 285 – 306. https://doi.org/10.14650/106228

Chai, T. (2020). Root Mean Square. In *Encyclopedia of Earth Sciences Series* (Vol. 2020). Springer Science and Business Media B.V. https://doi.org/10.1007/978-3-030-26050-7_280-1

Chiwuzie, A., Dabara, D. I., & Omotehinshe, O. J. (2021). Changing macroeconomic indicators and the rental values of residential properties in. *Article in YBL Journal of Built Environment*. https://doi.org/10.33796/ajober.5.1.01

Dwivedi, R., Gupta, R., & Pal, P. K. (2022). House Price Prediction using regression techniques. In S. V., S. V., S. M., S. M., & S. J. (Eds.), *Proceedings - 2022 4th International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2022* (pp. 165 – 170). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICAC3N56670.2022.10074128

Edalatpanah, S. A., Hassani, F. S., Smarandache, F., Sorourkhah, A., Pamucar, D., & Cui, B. (2024). A hybrid time series forecasting method based on neutrosophic logic with applications in financial issues. *Engineering Applications of Artificial Intelligence*, *129*. https://doi.org/10.1016/j.engappai.2023.107531

Eny, E. P., Nimran, U., Suharyono, & Kusumawati, A. (2020). Novelty Seeking to Predict Tourist Behavior Intention In Rural Destination. *Jurnal Bisnis Dan Manajemen*, *7*(2), 61–73. http://jurnal.unmer.ac.id/index.php/jbm

Forys, I. (2022). Machine learning in house price analysis: regression models versus neural networks. *Procedia Computer Science*, *207*, 435–445. https://doi.org/10.1016/j.procs.2022.09.078

Gandhi, B., Jaitly, N., Vats, K., & Sagar, P. (2023). A Prediction Model for Real Estate Price Using Machine Learning. *Lecture Notes in Networks and Systems*, *641 LNNS*, 37 – 42. https://doi.org/10.1007/978-981-99-0483-9_4

Geerts, M., vanden Broucke, S., & De Weerdt, J. (2023). A Survey of Methods and Input Data Types for House Price Prediction. In *ISPRS International Journal of Geo-Information* (Vol. 12, Issue 5). MDPI. https://doi.org/10.3390/ijgi12050200

Hamami, F., & Dahlan, I. A. (2024). Regression Modeling for House Price Prediction in Java Island. *ICSINTESA 2024 - 2024 4th International Conference of Science and Information Technology in Smart Administration: The Collaboration of Smart Technology and Good Governance for Sustainable Development Goals*, 402 – 406. https://doi.org/10.1109/ICSINTESA62455.2024.10747910

Hoxha, V. (2024). Comparative analysis of machine learning models in predicting housing prices: a case study of Prishtina's real estate market. *International Journal of Housing Markets and Analysis*. https://doi.org/10.1108/IJHMA-09-2023-0120

Jafari, A., & Akhavian, R. (2019). Driving forces for the US residential housing price: a predictive analysis. *Built Environment Project and Asset Management*, *9*(4), 515–529. https://doi.org/10.1108/BEPAM-07-2018-0100

Jáuregui-Velarde, R., Andrade-Arenas, L., Celis, D. H., Dávila-Morán, R. C., & Cabanillas-Carbonell, M. (2023). Web Application with Machine Learning for House Price Prediction. *International Journal of Interactive Mobile Technologies*, *17*(23), 85–104. https://doi.org/10.3991/IJIM.V17I23.38073

Kang, C.-D. (2025). Effects of Access Time via Travel Modes to Urban Service Facilities on Housing Prices: Evidence from Seoul. *Journal of Urban Planning and Development*, *151*(2). https://doi.org/10.1061/JUPDDM.UPENG-5513

Kavitha, S., Varuna, S., & Ramya, R. (2016). Proceedings of 2016 Online International Conference on Green Engineering and Technologies (IC-GET) : 19th November, 2016. *Proceedings of 2016 Online International Conference on Green Engineering and Technologies (IC-GET)*.

Kaytan, M., & Aydilek, I. B. (2017). IDAP'17 : International Artificial Intelligence and Data Processing Symposium : September 16-17. *IDAP'17 : International Artificial Intelligence and Data Processing Symposium*.

Kotu, V., & Deshpande, B. (2019). *Data Science: Concepts and Practice, Second Edition* (Second Edition). Morgan Kaufmann Publishers.

Krämer, B., Nagl, C., Stang, M., & Schäfers, W. (2023). Explainable AI in a Real Estate Context – Exploring the Determinants of Residential Real Estate Values. *Journal of Housing Research*, *32*, 1–42. https://doi.org/10.1080/10527001.2023.2170769

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In *Nature* (Vol. 521, Issue 7553, pp. 436–444). Nature Publishing Group. https://doi.org/10.1038/nature14539

Lenaers, I., Boudt, K., & De Moor, L. (2024). Predictability of Belgian residential real estate rents using tree-based ML models and IML techniques. *International Journal of Housing Markets and Analysis*, *17*(1), 96–113. https://doi.org/10.1108/IJHMA-11-2022-0172

Mishra, R., Singh, R., & Govindan, K. (2022). Net-zero economy research in the field of supply chain management: a systematic literature review and future research agenda. In … *Journal of Logistics Management*. emerald.com. https://doi.org/10.1108/IJLM-01-2022-0016

Mukhlishin, M. F., Saputra, R., & Wibowo, A. (2017). Predicting house sale price using fuzzy logic, Artificial Neural Network and K-Nearest Neighbor. *Proceedings - 2017 1st International Conference on Informatics and Computational Sciences, ICICoS 2017*, *2018-January*, 171 – 176. https://doi.org/10.1109/ICICOS.2017.8276357

Njo, A., Sugiarto, H. D., & Prayogo, D. (2025). Artificial intelligence algorithms to predict housing market prices in Surabaya, Indonesia. *International Journal of Housing Markets and Analysis*. https://doi.org/10.1108/IJHMA-01-2025-0022

Rahmatillah, I., & Sudirman, I. D. (2023). A Deep Learning Approach to Predict Housing Market Price Trends in Bandung. *2023 1st IEEE International Conference on Smart Technology: Advances in Smart Technology for Sustainable Well-Being, ICE-SMARTec 2023*, 140 – 145. https://doi.org/10.1109/ICE-SMARTECH59237.2023.10461965

Rampini, L., & Re Cecconi, F. (2021). Artificial intelligence algorithms to predict Italian real estate market prices. *Journal of Property Investment and Finance*. https://doi.org/10.1108/JPIF-08-2021-0073

Sabrina Abdul Latif, N., Muhammad Rizwan Majeed, K., Rozzani, N., & Khadijah Saleh, S. (2020). Factors Affecting Housing Prices in Malaysia: A Literature Review. *International Journal of Asian Social Science*, *10*(1), 63–67. https://doi.org/10.18488/journal.1.2020.101.63.67

Sanyal, S., Kumar Biswas, S., Das, D., Chakraborty, M., & Purkayastha, B. (2022). Boston House Price Prediction Using Regression Models. *2022 2nd International Conference on Intelligent Technologies, CONIT 2022*. https://doi.org/10.1109/CONIT55038.2022.9848309

Saraswathi, P., Prabha, M., Jhoshitha, N. A., & Prathiksaa, J. G. (2024). House price prediction with advanced machine learning techniques. In M. N.K. & R. S. (Eds.), *AIP Conference Proceedings* (Vol. 3214, Issue 1). American Institute of Physics. https://doi.org/10.1063/5.0239036

Statista. (2023). *Real estate in Indonesia*.

Statista. (2024). *Residential real estate in Indonesia*.

Szumilo, N., Laszkiewicz, E., & Fuerst, F. (2017). The spatial impact of employment centres on housing markets. *Spatial Economic Analysis*, *12*(4), 472 – 491. https://doi.org/10.1080/17421772.2017.1339119

Teoh, E. Z., Yau, W. C., Ong, T. S., & Connie, T. (2023). Explainable housing price prediction with determinant analysis. *International Journal of Housing Markets and Analysis*, *16*(5), 1021–1045. https://doi.org/10.1108/IJHMA-02-2022-0025

Trang, L. H., Huy, T. D., & Le, A. N. (2021). Clustering helps to improve price prediction in online booking systems. *International Journal of Web Information Systems*, *17*(1), 45–53. https://doi.org/10.1108/IJWIS-11-2020-0065

van Ruijven, K., & Tijm, J. (2024). Do People Value Environmental Goods? Evidence From the Dutch Housing Market. *Journal of Real Estate Literature*, *32*(2), 57–102. https://doi.org/10.1080/09277544.2024.2322781

Verma, A., Gera, P., Singhal, S., & Mohapatra, A. K. (2023). Advanced Regression Models for Accurate House Price Prediction: An Analysis of Performance and Interpretability. *2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023*. https://doi.org/10.1109/ICCCNT56998.2023.10307946

Wang, F., & Ran, G. (2019). Excessive Financial Support, Real Estate Development and Macroeconomic Growth: Evidence from China. *Emerging Markets Finance and Trade*, *55*(11), 2437–2447. https://doi.org/10.1080/1540496X.2018.1555463

Wang, Q., Lin, B., & Tan, C. (2024). A modified spatial house price to income ratio and housing affordability drivers study: using the post-LASSO approach. *International Journal of Housing Markets and Analysis*. https://doi.org/10.1108/IJHMA-12-2023-0169

Warisse, C. (2017). *Economic Review - June 2017 - Analysis of the developments in residential property prices : Is the Belgian market overvalued ?*

Whelan, S., Atalay, K., Barrett, G., Cigdem-Bayram, M., & Edwards, R. (2023). Transitions into home ownership: a quantitative assessment. *AHURI Final Report*, *404*, 1–96. https://doi.org/10.18408/ahuri7327101

Wu, H., Jiao, H., Yu, Y., Li, Z., Peng, Z., Liu, L., & Zeng, Z. (2018). Influence factors and regression model of urban housing prices based on internet open access data. *Sustainability (Switzerland)*, *10*(5). https://doi.org/10.3390/su10051676

Zhang, H., Li, Y., & Branco, P. (2023). Describe the house and i will tell you the price: House price prediction with textual description data. *Natural Language Engineering*. https://doi.org/10.1017/S1351324923000360

Zhang, Y., Rahman, A., & Miller, E. (2023). Longitudinal modelling of housing prices with machine learning and temporal regression. *International Journal of Housing Markets and Analysis*, *16*(4), 693–715. https://doi.org/10.1108/IJHMA-02-2022-0033

Zhao, J., & Zhang, B. (2021). Application and Study of artificial neural network intelligent control. *Journal of Physics: Conference Series*, *1948*(1). https://doi.org/10.1088/1742-6596/1948/1/012003

Zhao, Y., Chetty, G., & Tran, D. (2019). *2019 IEEE Symposium Series on Computational Intelligence : IEEE SSCI 2019 : December 6-9, 2019, Xiamen, China*. 1396–1401.

Zheng, Y. (2023). Community resilience and house prices: A machine learning approach. *Finance Research Letters*, *58*. https://doi.org/10.1016/j.frl.2023.104400

Zulkifley, N. H., Rahman, S. A., Ubaidullah, N. H., & Ibrahim, I. (2020). House price prediction using a machine learning model: A survey of literature. *International Journal of Modern Education and Computer Science*, *12*(6), 46–54. https://doi.org/10.5815/ijmecs.2020.06.04