PREDICTION OF RESIDENTIAL REAL ESTATE SELLING PRICES IN SERBIA USING ARTIFICAL NEURAL NETWORKS

DOI
https://doi.org/
10.18690/um.epf.5.2024.8

ISBN
978-961-286-867-3

LJILJANA MATIĆ,[1] VESNA RANKOVIĆ,[2] TIJANA GEROSKI,[2] ZORAN KALINIĆ[1]

[1] University of Kragujevac, Faculty of Economics, Kragujevac, Serbia
ljiljana.matic@ef.kg.ac.rs, zkalinic@kg.ac.rs

[2] University of Kragujevac, Faculty of Engineering, Kragujevac, Serbia
vesnar@kg.ac.rs, tijanas@kg.ac.rs

The increasing housing prices over the past decades have added complexity to the real estate appraisal process. Therefore, it is important to create a proper prediction model which can encapsulate complex dependence of the property price from variable inputs. But, the problem of predicting real estate prices is highly non-linear and depends on many parameters. This research explores the potential of utilizing artificial neural networks (ANNs) to forecast the selling prices of apartments in Belgrade, Serbia, based on various apartment parameters. The findings demonstrate high efficiency of the ANN models in property valuation and, if all the preconditions of property value modelling are met, the ANN technique stands as a reliable valuation approach that could be used by both real estate researchers and professionals.

Keywords:
real estate,
real estate market,
artificial intelligence,
artificial neural network,
price prediction

JEL:
R31,
R21,
C45

University of Maribor Press

1 Introduction

The real estate market is a crucial part of the global economy, affecting macroeconomic stability and social welfare. Understanding its dynamics is essential for academics and industry professionals. Analyzing property values and forecasting market trends are complex tasks influenced by numerous variables. Artificial intelligence (AI) offers advanced models for assessment and analysis, providing rapid and accurate insights.

Clients value the expertise of real estate professionals, but manual analysis is time-consuming. AI can quickly process vast datasets, offering valuable predictions on market trends and future investments. A Recent research in the Republic of Serbia has highlighted AI's practical application in real estate analysis, using data from a reputable Serbian real estate website.

This introduction sets the stage for exploring how AI and artificial neural networks revolutionize real estate analysis, providing investors with timely, data-driven insights for informed decision-making.

2 Literature review

Neural networks have played a significant role in the field of real estate valuation, offering powerful tools for predicting property prices.

Research using artificial neural networks dates back to the 1990s and has shown varied results. Early studies, such as those by Borst (1991), Do and Grudnitski (1992), Tay and Ho (1992), Worzala et al. (1995), and McCluskey (1996), found ANNs to be useful for property valuation. These studies compared ANN models with traditional statistical models, and often found artificial neural networks to outperform the latter. For example, Do and Grudnitski (1992) in the United States concluded that the neural network model performed better than the multiple regression model in estimating the values of residential properties. They used eight independent variables and found that their neural network model resulted in almost twice the number of predicted values within 5% of the actual sale price, compared to the multiple regression model (40% versus 20%), on a test set of 105 houses. The mean absolute error of their neural network model was significantly lower than the mean absolute error of their regression model (6.9% versus 11.3%). Similarly, studies by Nguyen and Cripps (2001) in the United States, Limsombunchai et al. (2004) in New Zealand, and Peterson and Flanagan (2009) in the United States supported earlier research claiming that neural networks provide better results than hedonic regression models. However, some studies have disagreed with these findings. Worzala et al. (1995) and McGreal et al. (1998) found that other models, such as additive regression (AR), M5P trees, and support vector machines (SVM), provided better results than neural networks. In mass appraisal, Lin and Mohan (2011) found that neural networks performed better than hedonic regression. Despite these differing results, artificial neural networks continue to be a valuable tool in real estate price prediction. Recent studies, such as those by Abidoye and Chan (2017), Aydemir et al. (2020), and Yılmazel et al. (2018), have shown promising results using neural networks for real estate price prediction.

In conclusion, while there is ongoing debate about the effectiveness of neural networks compared to the traditional methods and models in real estate valuation, they remain a widely used and promising tool in the field.

Further research and advancements in artificial neural network technology are likely to lead to even more accurate and reliable models for real estate value prediction in the future.

3 Methodology

Artificial neural networks are inspired by biological neural networks, representing their mathematical or computational model. Neurons are represented as processing elements and synapses as the weight of the connection. Dendrites are inputs and the axon is the output of the processing element. Processing elements are connected in a network such that the output of each is connected to the input of at least one other. The processing performed in the neuron's body is represented by input and transfer functions. The structure of the artificial neuron is shown in the Figure 1.

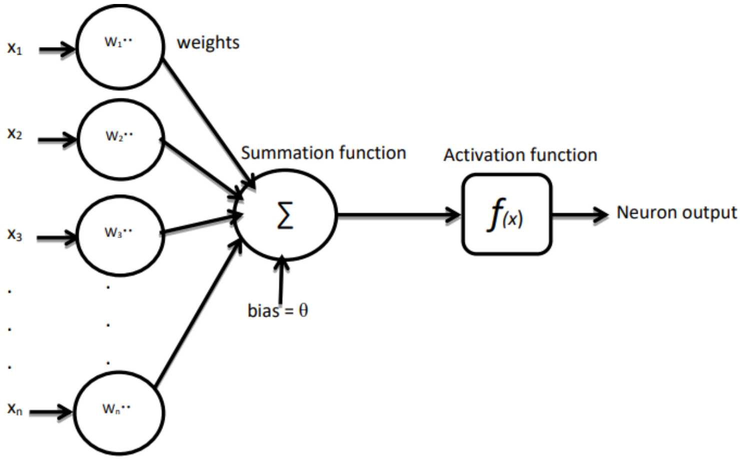

Figure 1: The structure of the artificial neuron

ANN models were proved to be very efficient in modelling complex, non-linear problems with many inputs. One drawback of this method is that it requires large datasets for training and testing of the model.

3.1 The data

The data was collected from one of the biggest and most significant online sales portals in Serbia, which deals, among other things, with real estate sales in the Republic of Serbia. Dataset was collected in period of December 2023-January 2024 and it consists of 8.358 active apartment sale listings in Belgrade, published in the previous four years. Table 1 lists all the collected parameters of each real estate unit i.e. apartment. The price parameter represents the asked selling price of the apartment. *Area r*epresents the total area of the apartment in square meters. *Rooms* is the total number of rooms in the apartment. *Bathrooms* is the total number of bathrooms in the apartment. *Floor* represents the floor at which the apartment is located and *Total Floors* is the total number of floors in the building. *Construction type* represents the physical condition of the apartment (Original condition, Renovated, Luxury, For renovation). *Date* is the year when the listing was posted, covering the period from 2020 to 2024. *Heating* is the type of heating in the apartment (Central, Electric heating, Heat pumps, Gas, Floor heating, Norwegian radiators, Marble radiators, A/C unit, Central heating with calorimeter, Tile stove, Solid fuel stove). *District* is the municipality where the apartment is located and *Settlement* is its microlocation within the municipality. The district and settlement parameters are combined into a new parameter, *Address*.

Table 1: The real estate parameters

Parameter	Data type	Parameter	Data type
Price	Currency (in €)	Construction type	String
Area	Number (in m2)	Publishing year	Number
Rooms	Number	Heating	String
Bathrooms	Number	District	String
Floor	Nmber	Settlement	String
Total Floors	Number	Address	String

Considering that the initial dataset lacked sufficient information for certain neighborhoods/settlements in Belgrade, only the data for neighborhoods with a minimum of 50 occurrences (for the larger model) and at least 150 occurrences (for the smaller model) in the dataset were used in the further model creation process. All duplicates and missing values were removed. Textual data (address, heating, construction type) underwent label encoding. All data were normalized using Min-Max normalization.

4 Results

The primary measure of loss is the absolute percentage error of prediction, where 0% represents a perfect prediction. The first model architecture consists of six layers. The first and second layers have a learning rate of 0.005, they use ReLU activation functions, have batch sizes of 64 and nodes per layer of 1024 and 512, respectively. The dropout rates for these layers are 0.2 and 0.3, respectively. For the third and fourth layers, the learning rate is also 0.005, with ReLU activation functions, batch sizes of 64, and nodes per layer of 256 and 128, respectively. The dropout rates for these layers are 0.4 and 0.5, respectively. The fifth layer has a learning rate of 0.005, uses LeakyReLU as the activation function, a batch size of 64, 64 nodes per layer, and a dropout rate of 0.5. Finally, the sixth layer has a learning rate of 0.01, uses a linear activation function, a batch size of 64, 32 nodes per layer, and no dropout. With this architecture, a loss of 19.97% was achieved. The algorithm was then applied to a smaller dataset, resulting in a 6.63% lower error rate, amounting to 13.3%.

On a smaller dataset, the best results are obtained with the momentum, when the value of the parameter beta_1 is 0.5, so this was the value used in the further process. The error was reduced to 12.3%.

According to the model by Çılgın and Gökçen (2023), a different model was created. A neural network was modeled with one input, three hidden, and one output layer. The input layer has 8 neurons with a tanh activation function.

The following three hidden layers each have 720 neurons and a ReLU activation function. Between each pair of layers, a dropout layer with a rate of 20% is placed. The output layer has one neuron with a ReLU activation function. The model was compiled using the Adam optimizer with a learning rate of 0.0015 and a mean squared error is used as the loss function. Two beta parameters were experimented with, beta_1=0.9 and beta_2=0.999. The error of this model is about 3% lower than the first model on a larger dataset, amounting to 17.59%.

In general, our research has shown that predictive models based on artificial neural networks have a satisfactory level of accuracy. Specifically, the predictive model for estimating real estate prices in the Serbian market, developed for this study, has an average deviation of the actual price from the forecasted price of up to 12%.

5 Discussion

Compared to Abidoye and Chan's (2017) neural network with one hidden layer and a mean absolute error of 16%, our neural network performed better, with an error of 12%. However, our model performed worse than the one by Del Giudice et al. (2017) with a mean absolute error below 7%. One reason for our model's higher error could be the actual property prices, influenced by factors beyond our analysis. Neural networks may also be inadequately trained or have inappropriate hyper-parameter values. Predicting prices in models with more data was challenging, possibly due to the dataset containing apartments from different areas with a wide price range. While mean, median, and standard deviation were similar for both datasets, the minimum and maximum values differed significantly.

The Improvement was seen with a smaller dataset. There may be an issue with the number of hidden nodes relative to the available data for training. While certain rules govern this relationship, they were not strictly followed in our research, as the optimal architecture depends on the problem type and techniques used.

The results are acceptable, but further exploration is needed to minimize the maximum error. Predicting price ranges instead of exact prices could be a solution. Comparing the multilayer perceptron with other architectures could provide additional insights. Expanding the research to include sales data from other areas could also be beneficial. Using separate models for each area might improve performance when dealing with large datasets containing diverse addresses.

6 Conclusions

The study examined the potential of predicting real estate sales prices in Belgrade, Serbia, using artificial neural networks. Following the research of Abidoye and Chan (2017), Varma et al. (2018), Aydemir et al. (2020), among many others, a neural network model was created. Neural networks resulted in performance differences depending on the size of the dataset, where the smaller dataset had a lower error rate. The smallest model error obtained on the larger dataset was approximately 17%, while on the smaller dataset, it was approximately 12%. Compared to the results of Abidoye and Chan (2017) who achieved a mean absolute error of 16%, the neural network showed better performance, with an error of 12%. However, compared to the model created by Del Giudice et al. (2017), with a mean absolute error below 7%, our model performed worse. Overall, the models behaved as expected and achieved similar results to other studies on the same topic. In future research, it is possible to compare different and current approaches with the methods used in this study, as well as test hyper-parameters with wider intervals in the methods used in the study. Also, with slight modifications and different dataset, the same model can be used to predict the rental prices of the apartments.

The Presented research has several limitations, all of which are potential avenues for further research. First, the developed prediction model in current form is usable for apartments as just one of the forms of real estate. It would be necessary to modify set of inputs and the choice of hyper-parameters, in order to apply it to other types of the real estate, such as houses, land or commercial properties. Next, the presented research was focused on Belgrade. It would be interesting to apply the same methodology and model on different cities/countries. One of the main limitations is the fact that the selling prices asked by owners in the listings were used as estimated values of the apartments, which could be biased, since owner's wishes are usually unrealistic and over-optimistic i.e. the real values of the apartments are usually lower than the asked prices in the listings. One way to resolve this issue is to use data for officially registered apartment sales by contracts from Tax Administration Office, but at this moment this kind of data is very hard to get, at least to the extent necessary for the application of ANN models. Finally, ANN models might not be the most appropriate technique for this problem, since there are many categorical model inputs. It would be interesting to apply some other machine learning techniques, such as Random Forest or XGBoost, and compare the obtained results.

References

Abidoye, R. B., & Chan, A. P. C. (2017). Modelling property values in Nigeria using artificial neural network. *Journal of Property Research*, 34(1), 36–53.

Aydemir, E., Aktürk, C., & Yalçınkaya, M. A. (2020). Estimation of Housing Prices with Artificial Intelligence. *Turkish Studies*, 15(2), 183–194.

Borst, R. A. (1991). Artificial neural networks: The next modelling/calibration technology for the assessment community. *Property Tax Journal*, 10(1), 69–94.

Borst, R. A. (1995). Artificial neural networks in mass appraisal. *Journal of Property Tax Assessment & Administration*, 1(2), 5–15.

Çılgın, C., & Gökçen, H. (2023). Machine Learning Methods for Prediction Real Estate Sales Prices in Turkey. *Revista de la Construcción. Journal of Construction*, 22(1), 163–177. [DOI: 10.7764/RDLC.22.1.163]

Del Giudice, V., De Paola, P., & Forte, F. (2017). Bayesian Neural Network Models in the Appraisal of Real Estate Properties. *In Computational Science and Its Applications – ICCSA*, 478–489.

Do, A. Q., & Grudnitski, G. (1992). A neural network approach to residential property appraisal. *The Real Estate Appraiser*, 58(3), 38–45.

Limsombunchai, V., Gan, C., & Lee, M. (2004). House price prediction: Hedonic price model vs artificial neural network. *American Journal of Applied Sciences*, 1(3), 193–201.

Lin, C. C., & Mohan, S. B. (2011). Effectiveness comparison of the residential property mass appraisal methodologies in the USA. *International Journal of Housing Markets and Analysis*, 4(3), 224–243.

McCluskey, W. M. (1996). Predictive accuracy of machine learning models for the mass appraisal of residential property. *New Zealand Valuers' Journal*, July, 41–47.

McGreal, S., Adair, A., McBurney, D., & Patterson, D. (1998). Neural networks: The prediction of residential values. *Journal of Property Valuation and Investment*, 10(1), 57–70.

Nguyen, N., & Cripps, A. (2001). Predicting housing values: A comparison of multiple regression analysis and artificial neural networks. *Journal of Real Estate Research*, 22(3), 313–336.

Peterson, S., & Flanagan, A. B. (2009). Neural network hedonic pricing models in mass real estate appraisal. *Journal of Real Estate Research*, 31(2), 147–164.

Tay, D. P. H., & Ho, D. K. K. (1992). Artificial intelligence and the mass appraisal of residential apartment. *Journal of Property Valuation and Investment*, 10, 525–540.

Varma, A., Sarma, A., Doshi, S., & Nair, R. (2018). "House Price Prediction Using Machine Learning and Neural Networks". In 2018 *Second International Conference on Inventive Communication and Computational Technologies* (ICICCT), 1936-1939. IEEE.

Worzala, E. M., Lenk, M. M., & Silva, A. (1995). An exploration of neural networks and its application to real estate valuation. *Journal of Real Estate Research*, 10(2), 185–202.

Yılmazel, Ö., Afşar, A., & Yılmazel, S. (2018). Using Artificial Neural Network Method to Predict Housing Prices. *International Journal of Economic & Administrative Studies*, (20), 285–300.

Zdravković, S., Pavlović, G., & Peković, J. (2019). Determinants of consumer intentions regarding future use of mobile commerce - the moderating effect of personal innovativeness. *Marketing*, 50(2), 124–134.