Real Estate in Vilnius Price Prediction Using Artificial Neural Network Model

Justina Gertaitė, Mantas Micikevičius, Rokas Steponavičius

Abstract

Today the real estate market has become very popular. According to Ober-Haus Real Estate Advisors the prices of apartments in Vilnius in September 2021 increased by 19.9% on a year-over-year basis and the price per square meter reached a record after almost 14 years [1]. As real estate prices in Vilnius increase, it is difficult for housing purchasers to identify if the price of an apartment or a house is overpriced or reasonable to buy. Moreover, a reasonable price prediction of the target property will add effectiveness for residential developers, as they could decide whether the expected return from a housing development project is sufficient to offset the risks. Unfortunately, there is no free access to find historical data to analyze and predict the real estate in Vilnius appraisal compared to the market. Considering the problem, this paper aims to demonstrate how using data scraping method and data mining techniques it is possible to detect if real estate prices in Vilnius are appropriate. In order to collect historical real estate market data in Vilnius a web scraping method using Python language is used. Model based on the Artificial Neural Network (ANN) is used for the price prediction part. As it is hard to find reliable information and analysis of the real estate market in Vilnius, the model in this study provides helpful information for residential developers and buyers in Vilnius. Nevertheless, the model could be implied for any other city.

Keywords: Artificial Neural Network (ANN), web scraping, real estate.

1. Literature review

In this paper web scraping will be used, which is data scraping. Data scraping is a technique where a computer program extracts data from human-readable output coming from another program and saves it to a file system for later analysis. When there is limited access to the data, data scraping is one of possible ways to succeed [2].

João Manuel Azevedo Santos (2018) in his dissertation to analyse the real estate market for financial investments proposes the development of a web scraping application for collecting real estate data. In the mentioned dissertation were chosen real estate advertising web platforms in Porto district and a system was created to collect data over time and detect

changes in real estate market prices. In our case, due to limited time, we will not use data extraction over time, only once. Furthermore, in the dissertation a model was created using classification, random forest and single layer perceptron to predict real estate asset's price fluctuations [3].

Artificial Neural Network (ANN) is an artificial intelligence model that replicates the human brain's learning process. Artificial Neural Networks are able to learn, to generalize results and to respond adequately to highly incomplete or previously unknown data [4]. ANN methodology was developed to capture functional forms, allowing the uncovering of hidden non-linear relationships between the variables. One of the greatest advantages of artificial neural networks is their capability to learn from their environment. Learning from the environment comes useful in applications where complexity of the environment (data or task) make implementations of other types of solutions impractical [5]. It has been applied to property price forecasting in recent years. There is a lot of research related to this paper.

In related scientific paper to our topic Vincenza Chiarazzo, Leonardo Caggiani, Mario Marinellia and Michele Ottomanellia (2014) applied traditional multiple regression models and the estimated neural network models in showing how different transport characteristics and the environmental quality affect the prices of real estate properties [4].

Itedal Sabri Hashim Bahia (2013) applied two different Artificial Neural Network techniques (Feed forward back propagation, FFBP and cascade forward back propagation, CFBP) to predict housing prices and compare the result between them. CFBP Network prediction was more accurate on selected parameters [6].

Lai Pi-ying (2011) found out that using ANN is able to overcome multiple regressions' methodological problems and also get better performance than multiple regression models (MRA). This research investigated the problem of real estate appraisal in Kaohsiung city [7].

Selim (2011) examined the determinants of house prices in Turkey. Two types of modelling approaches were employed in the analysis: hedonic regression model and ANN. The results of the hedonic model revealed that water system, pool, type of house, number of rooms, house size, locational characteristic and type of the building are the most significant variables that affect the house prices. Because of potential non-linearity in the hedonic functions, ANN is employed as an alternative method for the prediction. By comparing the prediction performance between the hedonic regression and ANN models, this study demonstrates that ANN is a more accurate method for prediction of the house prices in Turkey [8].

Peterson and Flanagan (2009) using a large sample of 46,000 residential property sales spanning 1999–2005 period, demonstrate that, relative to linear hedonic pricing models, ANN generate significantly lower pricing errors, have greater pricing precision and extrapolate better from more volatile pricing environments. While a single layer ANN is functionally equivalent to OLS, multiple layered ANNs are capable of modelling complex non-linearities [9].

Nguyen and Cripps (2001) compared the predictive performance of artificial neural networks (ANN) and multiple regression analysis (MRA) for housing sales. Multiple comparisons were made between the two data models in which the data sample size, the functional specification and the temporal prediction were varied. ANN performed better than MRA when a moderate to large data sample size is used. The MRA performed better than the ANN when a small data sample size is used [10].

In the empirical study (Lam et al., 2008), entropy and ANN were integrated to forecast the housing price in Hong Kong. With the purpose of exploring an optimal forecasting model, different tests were carried out to explore the capability of ANN in handling different sample sizes. In the research, entropy was firstly applied to prioritize the price determinants after summarizing the microeconomic and macroeconomic variables which affect the housing price. Second, ANN is proposed to predict the property price.The overall results of this research demonstrated that the ANN serves a more desirable function of housing price forecasting with suitable input variables and relatively smaller sample sizes [11].

In this paper, an ANN approach is proposed with an analysis of performances in estimating the sale price of residential properties depending on factors.

References

[1] Ober-Haus Real Estate Advisors. (2021). Accessed 16 October 2021. `http://ober-haus.com/the-new-apartment-price-record-in-vilnius-after-almost-14-years/`.

[2] Zhao, Bo. (2017). Web Scraping. 10.1007/978-3-319-32001-4_483-1.

[3] João Manuel Azevedo Santos. (2018). Real Estate Market Data Scraping and Analysis for Financial Investments. `https://repositorio-aberto.up.pt/bitstream/10216/116510/2/296684.pdf`.

[4] Ottomanelli, Michele & Chiarazzo, Vincenza & Marinelli, Mario & Caggiani, Leonardo. (2014). A Neural Network based Model for Real Estate Price Estimation Considering Environmental Quality of Property Location. Transportation Research Procedia. 3. 810-817. 10.1016/j.trpro.2014.10.067.

[5] Krenker, A., Bešter, J., & Kos, A. (2011). Introduction to the artificial neural networks. Artificial Neural Networks: Methodological Advances and Biomedical Applications. InTech, 1-18.

[6] I. Bahia. (2013). A Data Mining Model by Using ANN for Predicting Real Estate Market: Comparative Study. International Journal of Intelligence Science, Vol. 3 No. 4, pp. 162-169. 10.4236/ijis.2013.34017.

[7] Lai, Peddy. (2011). Analysis of the Mass Appraisal Model by Using Artificial Neural Network in Kaohsiung City. Journal of Modern Accounting and Auditing. Vol. 7.

[8] Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. Expert systems with Applications, 36(2), 2843-2852.

[9] Peterson, S., & Flanagan, A. (2009). Neural network hedonic pricing models in mass real estate appraisal. Journal of real estate research, 31(2), 147-164.

[10] Nguyen, N. and A. Cripps. Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks. Journal of Real Estate Research, 2001, 22:3, 313–36.

[11] Lam, K. C., Yu, C. Y., & Lam, K. Y. (2008). An artificial neural network and entropy model for residential property price forecasting in Hong Kong. Journal of Property Research, 25(4), 321-342.