HYBRID PREDICTIONS OF THE HOMOGENOUS PROPERTIES' MARKET VALUE WITH THE USE OF ANN

H. ANYSZ[1], M. PODWÓRNA[2], N. IBADOV[3], K. LENNERTS[4], K. DIKAREV[5]

The homogenous properties – as flats are – have the set of key features that characterizes them. The area of a flat, the number of rooms and storey number where it is located, the technical state of a building, and the state of the vicinity of the blocks of flats assessed. The database comprises 222 flats with their transaction prices on the secondary estate market. The analysed flats are located in a certain quarter of Wrocław city in Poland. The database is large enough to apply machine learning for successful price predictions. Their close locations significantly lower the influence of clients' assessments of the attractiveness of the location on the flat's price. The hybrid approach is applied, where classifying precedes the solution of the regression problem. Dependently on the class of flats, the mean absolute percentage error achieved through the calculations presented in the article varies from 4,4 % to 7,8 %. In the classes of flats where the number of cases doesn't allow for machine predicting, multivariate linear regression is applied. The reliable use of machine learning tools has proved that the automated valuation of homogenous types of properties can produce price predictions with the error low enough for real applications.

Keywords: real estate valuation, ANN, comparative approach, machine learning, hybrid models

[1] PhD. Eng., Warsaw University of Technology, Faculty of Civil Engineering, Al. Armii Ludowej 16, 00-637 Warsaw, Poland, e-mail: h.anysz@il.pw.edu.pl

[2] PhD. Eng., Wrocław University of Science and Technology, Faculty of Civil Engineering , Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland, e-mail: monika.podworna@pwr.edu.pl

[3] PhD. DSc. Eng., Warsaw University of Technology, Faculty of Civil Engineering, Al. Armii Ludowej 16, 00-637 Warsaw, Poland, e-mail: n.ibadov@il.pw.edu.pl

[4] Prof., Karlsruhe Institute of Technology, Institute of Technology and Management in Construction, Gotthard-Franz-Street 3, 76131 Karlsruhe, Germany, e-mail: kunibert.lennerts@kit.edu

[5] PhD. Eng., Prydniprovska State Academy of Civil Engineering and Architecture, Department of Construction Technology, 24a, Chernyshevskogo St., Dnipro, 49005, Ukraine, e-mail: kdikarev@ukr.net

1. INTRODUCTION

Real estate valuation is a complex, multidisciplinary problem. The property appraiser must be fluent in economic, legal, and technical sciences. The expert should have these special abilities and experience, but he/she should be also fluent in accounting. It is an often case in property valuation when it is necessary to know the range of property values a short time. The situation is simpler if it concerns typical real estate, often marketed, e.g. average residential premises in a large city. Automated valuation models (AVM) are a controversial issue in both the scientific community and practitioners, but this doesn't mean that one should not look for new methods of valuation than specified in the regulations (they are presented in table 1). Due to the limited number of properties and much lower number of transactions on the real estate market, the most commonly used method is the comparative approach with pair comparative methods (a few similar real estate transactions are sufficient) or average price correction methods (several free-market similar real estate transactions are required then). The method of statistical analysis of the market is a more complicated and requires a much larger number of similar real estate transactions to analyse in the valuation process. The wide application of statistical methods and econometric models in real estate estimations are described e.g. in [1, 2].

Table 1. Approaches, methods and techniques of property valuation in Poland

Approaches	Methods	Techniques
Direct, comparative (market approach)	Pairwise comparative	----
	Adverse price correction	
	Statistical market analysis	
Income	Investment	Straight capitalization
		Discounted cashflow
	Profit	Straight capitalization
		Discounted cashflow
Cost	Replacement cost	Detailed
		Integrated elements
		Indicators
	Reconstruction cost	Detailed
		Integrated elements
		Indicators
Mixed Approach	Residual	----
	Land estimation indicators	
	Liquidation cost	

The comparative approach is recommended in the International Valuation Standards (IVS) [5] and in the European Valuation Standards (EVS) [4] as the main approach to receive market value of a property. The definition of the market values are:

- International Valuation Standards: "Market Value is the estimated amount for which an asset or liability should exchange on the valuation date between a willing buyer and a willing seller in an arm's length transaction, after proper marketing and where the parties had each acted knowledgeably, prudently and without compulsion." [5]
- European Valuation Standards: "The estimated amount for which the property should exchange on the date of valuation between a willing buyer and a willing seller in an arm's length transaction after proper marketing wherein the parties had each acted knowledgeably, prudently and without being under compulsion." [4]
- Polish Act on real estate management: The market value of the property is the estimated amount that can be obtained on the valuation day for the property in a sale transaction concluded on market terms between the buyer and the seller who are firmly willing to conclude the contract, act with insight and act prudently and are not in a forced situation [1].

"Market Value is a key concept in establishing an informed expectation as to the price for something, one that is neutral as between buyer and seller" [1]. The comparative approach estimates the market value for a property using recent sales data from other similar properties on the local market. This approach requires that there is an active market for similar properties. Also, local market conditions, as well as national economic conditions, should be stable. The approach considers the selling prices of similar, recently sold properties. Those sales prices are adjusted to reflect the time, conditions, and differences between the comparable properties and the evaluated property. The result of the adjustments is a subjective process. "To determine the comparative value, prices of properties are to be referred to which sufficiently correspond to the property to be valued in terms of the sustainable features that have a major effect on its value, in particular location, interiors and possible types of use" [1]. For determining the value of the real estate being assessed, it is assumed that the future buyer will pay for the real estate near the same price as the transaction prices obtained for similar real estates traded in the market, taking into account price adjustments for time-lapse and properties that differentiate the two specific real estates from each other.

Some researchers deal with mass valuation of properties which can be necessary for many purposes like property tax, price indices construction, understanding market dynamics, estimating investor's risks. The methods of mass valuation are collected e.g. in [6]. There is the review of the conventional Multiple Regression Analysis and more advanced methods such as Spatial Auto-Regressive model,

including Spatial Lag Model, Spatial Error Model, General Spatial Model, and local approach including Geographically Weighted Regression, Moving Window Regression, and geostatistical including Regression Kriging, and Moving Window Kriging. The multiple regression models may constitute a good tool for real estate valuation, but only in the conditions of a developed, well-functioning real estate market [7, 18, 19]. A Rough Set Theory with Value Tolerance Ratio is proposed for assessments made based on small markets [8]. Artificial neural networks (ANN) are also successfully applied in the valuation of construction works [23] (one of the methods of valuation of a building), but also in property valuations [17, 19, 24, 28, 29, 30].

2. PREDICTIONS OF THE PROPERTY'S MARKET VALUE

2.1. DATABASE

The type of properties considered in the analysis were flats in Wrocław city, in south-west Poland. There were 222 transactions made during the two years (2013-2014) found in the register by a real estate expert. Analysis carried out proved that there were no clear price changes within the considered period – neither increase nor decrease in the local flats market. To avoid the influence of the clients' preferences (concerning the quarter of Wrocław city) the record is limited to the transactions concerning flats in Grabiszyn (south-west part of Wrocław), in Fabryczna district. The types of data retrieved from the register are as follows: market price in PLN, area of a flat in m2, location of a flat – storey in a block of flats, number of rooms in a flat. The unit price of a square meter of a flat could be calculated. The basics statistics concerning these 222 flats are presented in Table 2.

Table 2. Basic statistics of 222 flat transactions

	price [PLN]	area [m²]	storey	rooms	unit price [PLN/m²]
minimum	62 000	15,96	0*	1	2 047,56
maximum	660 000	94,30	12	4	8 048,37
mean average	239 869	48,56	3,4	2,1	5 001,01
standard deviation	89 171	15,80	2,3	0,8	1 045,75
median value	227 500	47,93	3	2	4 852,10

* 0 means ground floor

Additionally, two other parameters are assessed for each flat i.e. technical state of a block of flats and the nearest vicinity of a block of flats considering the comfort of living. As the assessment of the technical state of blocks of flats is done based only on their external look and the look of entrance halls. The buildings are distinguished into two groups: moderate and a good state. The methods of assessing the technical state of a building are complex [10, 20, 21, 22]. A very simplified form of technical condition assessment was adopted for the article. Similarly, the nearest vicinity is assessed as a moderate or comfort one. A comfort vicinity was recognized when there were: green area, sport utilities, public communication stops, groceries located nearby. The independent variables applied in the property valuation were analysed based on [14, 18, 30]. Finally, the database consists of four numerical properties (price, storey, number of rooms, area of a flat) and two qualitative properties (technical state of a block of flats, the nearest vicinity) of 222 flats. Part of the database is presented in Table 3.

Table 3. Exemplary entries in the database

record number	price [PLN]	flat area [m²]	storey	number of rooms	technical state of a building	vicinity
1	184 000	36,85	4	2	Good	moderate
21	250 000	48,87	4	3	Good	comfortable
46	235 000	47,30	5	2	Good	comfortable
84	210 000	38,02	9	2	Good	moderate
137	234 000	41,35	2	2	Good	moderate
179	350 000	52,20	3	2	Moderate	comfortable
185	270 000	60,00	3	2	Good	comfortable
197	270 000	65,80	3	3	Good	moderate
202	415 000	76,33	3	3	Good	moderate
203	260 000	78,37	3	3	Good	comfortable
219	160 000	34,76	4	1	Good	moderate
222	160 000	36,70	4	2	Good	comfortable

Based on [11, 12, 13] linear standardization maximum is applied to four numerical types of data with the use of the following formula:

(2.1)
$$for \; 1 \leq i \leq k \quad a_{1i} = \frac{a_{0i}}{max_k(a_{0i})}$$

where:

a_{0i} – the value of i element before standardization, $max_k (a_{0i})$ – the maximum value of a certain type of data before standardization, a_{1i} – the value of a_{0i} after standardization, k – the number of rows in a database.

This made the value of all numerical data ranged from 0 to 1. This process is necessary to achieve good results when artificial neural networks (ANN) are applied. It enables ANN to "recognize" themselves which type of input influence the output the most, and which type of input can be neglected [11]. A suggestion made by an ANN user, by leaving original values (unit price value is a hundred times higher than the value of flat's area) brings usually poor results [11, 13].

2.2. PREDICTIONS

For the first attempt to a flat's unit price prediction with the use of ANN (MLP type; multilayer perceptron), the input consist of three types of numerical data (area of a flat, number of rooms, and the storey), and two types of qualitative data (vicinity of the building and its technical state). The price of a flat is set as an output. With the use of Statistica 13.1 software (by Dell) 5 the best ANNs are found. Their mean absolute percentage errors (MAPE) [13] are based on the following formula:

(2.2)
$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{v_i^{(p)} - v_i^{(o)}}{v_i^{(o)}} \right|}{n} * 100 \, \%$$

where:

$v_i^{(p)}$ – the predicted i value of an independent variable v, $v_i^{(0)}$ – observed i value of an independent variable v, n – number of predictions

The errors calculated for the validating dataset vary from 15,5 % to 16,6 % which is definitely too much – the predictions can't be useful [25]. The lowest MAPE is produced by ANN with 3 neurons in a hidden layer, exponential function as an activation function in the hidden layer, and linear function in the output layer. Even Figure 1, where standardized predictions and real values of the output for the validation dataset are presented, proves the weakness of this approach. The validation dataset consists of 15 % (33 out of 222) randomly chosen from all records.
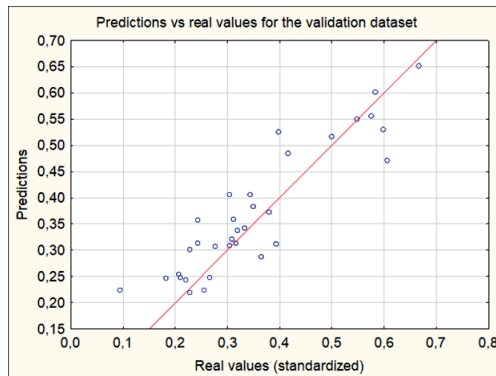
Fig. 1. Real prices vs predictions (both standardized) for the best predicting ANN

Based on the histogram of the unit prices (see Figure 2), it can be found that 64,0 % of the unit prices are from the range 4 000 - 6 000 PLN (142 out of 222 cases).
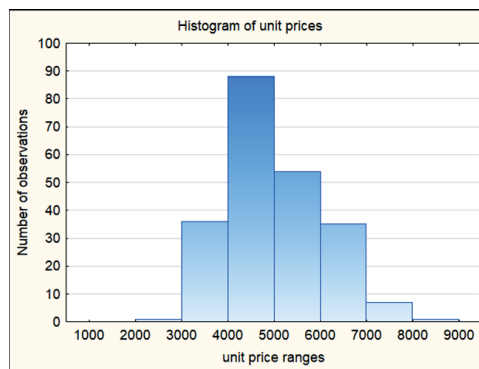

Fig. 2. The histogram of the flats' unit prices in PLN

It was a reason for searching the ANN, which could classify – as the first step – all cases to the three following classes with:

- low unit prices – below 4 000 PLN/m^2, (labelled L)
- middle ranged unit prices – from the range 4 000 – 6000 PLN/m^2, (labelled M)
- high unit prices – above 6 000 PLN/m^2 (labelled H).

The input to ANN remained unchanged. There are 3 ANNs found, which provide 100 % accuracy of classifications in all three sets of data (training, testing, and validating). The confusion matrix is presented in Table 4.

Table 4. Confusion matrix for the three best classifying ANNs

ANN type		Class L	Class M	Class H	All classes
MLP 7-16-3	Number of cases	37	142	43	222
	Correctly classified	37	142	43	222
	Incorrectly classified	0	0	0	0
MLP 7-20-3	Number of cases	37	142	43	222
	Correctly classified	37	142	43	222
	Incorrectly classified	0	0	0	0
MLP 7-13-3	Number of cases	37	142	43	222
	Correctly classified	37	142	43	222
	Incorrectly classified	0	0	0	0

All the above specified ANNs have hyperbolic tangent as the activation function in the hidden layer and softmax function [11] in the output layer. Moreover, several ANNs were producing only one wrongly classified case. So good classifying tool can be a base for introducing a two-stage (hybrid) tool for predicting prices of a flats (see Figure 3).
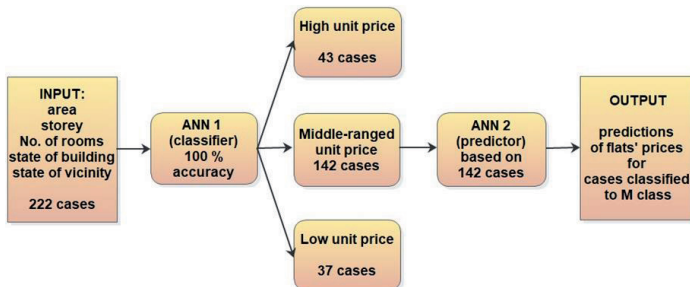

Fig. 3. Proposed the two-stage prediction tool for the cases classified to M class

Based on the successful use of ANN for a relatively low number of records in a database [27], even the records leading to class H or L are excluded from the database, 142 cases are sufficient to find the best predicting ANN 2 (as presented in Figure 4). Again, the types of input variables remained unchanged but the output is a flat's price. Within the five best predicting ANNs found, MAPE varies from 6,4 % to 7,0 %. In three of them, MAPE is below 6,5 %. The much lower error of predictions can be also observed in Figure 4 (compared to Figure 1), where predictions and their relevant real values are presented. The lowest MAPE is achieved for ANN with 9 neurons in a hidden layer and exponential activation function in the hidden and output layers.
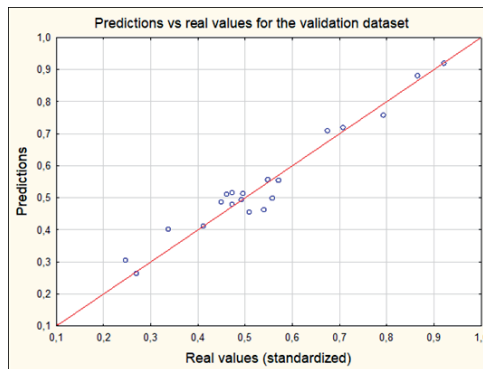

Fig. 4. Real prices vs predictions (both standardized) for the best predicting ANN 2
(for the middle-range unit prices database, for the validating dataset)

The mean absolute percentage error (MAPE), this time, is on a much more acceptable level. The level of MAPE for the training and testing datasets (considered as a one dataset) is slightly lower. It varies from 6,2 % to 6,9 %. For a new case, not included in the database, if it is classified to class M and next, e.g. 300 000 PLN price is predicted, MAPE= 6,5 % means, that the expected price should be in the range from 280 500 to 319 500 PLN. To have the predicting tool completed, the method for the two other classes (H and L) should be proposed. As both sets are too small to the effective application of ANN, and based on [14, 18, 19], multivariate, linear regression is proposed. As the flats' prices in these two groups vary much (e.g. 7 000 PLN/m2 is more than twice higher than 3 000 PLN/m2) preferences of clients can also be extremely varied. So, the linear regressions are searched separately for these two sets (H and L). They are based on the following formulas:

(2.3) $$v^{(L)} = x_1 * a_1^{(L)} + x_2 * a_2^{(L)} + x_3 * a_3^{(L)} + x_4 * a_4^{(L)} + a_6^{(L)}$$

(2.4) $$v^{(H)} = x_1 * a_1^{(H)} + x_2 * a_2^{(H)} + x_3 * a_3^{(H)} + x_4 * a_4^{(H)} + x_5 * a_5^{(H)} + a_6^{(H)}$$

where:

v – the predicted price of a flat, (L) – superscript meaning "for the cases classified as low unit price", (H) – is used for high unit price, x_1 – area of a flat. x_2 – the number of rooms, x_3 – the storey number, x_4 – vicinity of a block of flats defined in the formula (2.5), x_5 – technical state of a block of flats defined in the formula (2.6) below.

(2.5) $$x_4 = \begin{cases} 0 & for\ moderate \\ 1 & for\ comfortable \end{cases}$$

(2.6) $$x_5 = \begin{cases} 0 & for\ moderate \\ 1 & for\ good \end{cases}$$

For the dataset with low unit prices the independent variable x_5 is omitted, as all 37 records in the database have moderate technical state of the building, so they affect the independent variable with the same power, and it can be considered in $a_6^{(L)}$ value. To make the result of these two multivariate linear models comparable to the result for ANN 2, two sets (for H and L) of the constants a are searched for minimum MAPE in each model. The Statistica 13.1 (by Dell) software has produced much worse results than Microsoft Excel for Office 365 with Solver add-in (with nonlinear algorithm applied). The results i.e. two sets of the constants a and minimum MAPE errors are presented in Table 5.

Table 5. Results from Microsoft Excel Solver

Data set	Number of cases	a_1	a_2	a_3	a_4	a_5	a_6	Min MAPE
H	43	5 919,65	848,98	2515,76	6225,07	19 345,30	1 840,28	4,4 %
L	37	3 245,07	-785,18	10 074,56	-3 173,32	0	662,62	7,8 %

The much lower MAPE for H dataset is also visible in Figure 5. Predictions based on the multivariate linear model, created for the high unit prices (above 6 000 PLN/m2), are almost perfect when predicted price belongs to the range 100 000 – 275 000 PLN (see Figure 5 b).

b)

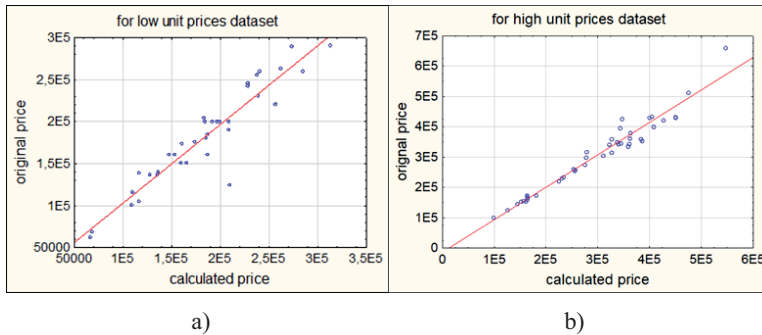a)   b)

Fig. 5. The comparatives of calculated and original flats' prices in PLN for a) L dataset, b) H dataset

For the prediction of a flat located in the specified quarter of Wrocław city, the invented, hybrid, two-stage model requires classification of the case (by ANN 1) to one of the three classes (L, M, H). It is done with 100 % accuracy by ANN 1 found above. Then – for the case belonging to L or H class – respective multivariate, linear function ($v^{(L)}$ or $v^{(H)}$) is used for prediction with MAPE 7,8 % for L, and 4,4 % for H. If ANN 1 classifies the case to M class, then ANN 2 should be used for flat's price prediction with MAPE equal to 6,5 %. The scheme of the completed model is presented in Figure 6.
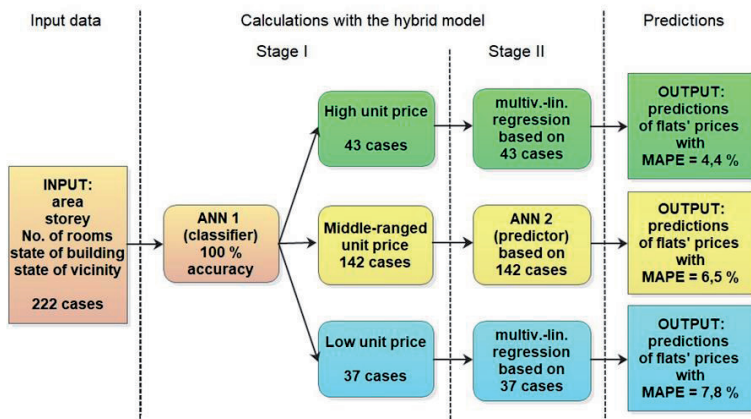
Fig. 6. The scheme of the invented model for flats' prices predicting

To evaluate the overall MAPE for the proposed predicting tool, the error can be weighted by the number of elements in each class (L-M-H). It can be calculated then, that overall MAPE for the tool is 6,3 %.

3. DISCUSSION

The process of property appraisal is regulated in Poland and many other states [4, 9]. There are also international standards for this business activity [3, 5]. Automated valuation of a building or any other construction object is questionable, as the sole base for the official appraisal report. The problem, if the proposed method will be accepted by law regulations, is left for practitioners, legal property appraisers. The prediction of the market value of secondary estate market flats – based on high accuracy achieved in the proposed model – can be certainly applied as an auxiliary tool. To do so, several conditions should be met.

The type of a property as flats are, considering all type of properties, are very homogenous. Homogeneity of flats is one of the crucial conditions. The same set of their features influence their price. This approach can be found also in [28, 30]. Assessing the value of an office building or a production hall with the model based on flats could give an unacceptable high error. Another issue, strongly connected with homogeneity is the number of transactions recorded in the secondary estate market. What is possible for flats (where the high number of transactions is observed), it can be impossible to build a similar model e.g. for wooden houses or steel masts, which are rather unique structures. The next problem is the variation in the prices over time. It is overcome in the article by considering the transactions that had a place only in 2013 and 2014, in the period of flat market stability. If the market is not stable, or the transactions are recorded in a much longer period, this factor should be represented in the input to the model [16, 17]. The proposed model doesn't assume any data concerning the location of a flat in a city. The homogeneity of assessed properties refers to the location too. The dataset comprises the flats located in the same quarter of Wrocław city, so the model could omit a location as a type of input. However, the clients' preferences and their perceptions of the locations of flats may influence the price much [14, 15]. Considering all conditions mentioned above, it can be stated the model of predicting secondary estate market flat prices is tailored to the database collected. The necessity of matching the prognostic model to the kind of properties assessed by it is confirmed by significantly increased accuracy of the hybrid (2 staged) model to the accuracy of the ANN model created for all 222 cases. The mean absolute percentage error of predicted flats' prices equals to 6,3 % can be evaluated as very low when compared to the scale of MAPE [25], as well as when compared to other predictions [24, 29, 30] of properties' value. The proposed model can be applied in the present form for valuating flats located in one district of a city, if the large enough set of data is provided. The flats' prices should be current, if the current prices are predicted.

Despite considered the limitations described above arisen also from market conditions, the successful application of the proposed model wouldn't be possible without 100 % accurate classification achieved in stage I of the hybrid model. Based on the histogram (presented in figure 2) and abilities of ANN the classification of flats' unit prices is recognized and applied. Perfect, automatic classification allows avoiding conditional percentage error discussion in stage II. Then the two approaches to the flats' prices prediction can be applied (ANN for the middle range unit prices group with a sufficient number of cases, and multivariate regression for H and L groups, where ANN couldn't be applied). Hybrid models – as presented in the article and comparing e.g. results of [26] and [27] – can lower prediction errors as more than a single tool is engaged in prediction processes.

4. CONCLUSIONS

Market phenomena influencing secondary estate market flats' prices are modelled with the use of artificial neural networks and multivariate linear regression. The two-stage hybrid model is proposed. The set of transactions is classified into three groups with different ranges of the unit prices as a first step. It is made with 100 % accuracy. The ANN model is built for flats' prices predictions for the most numerous group of flats. The mean absolute percentage error is 6,5 % there. For the groups of low and high unit prices separate models are created based on multivariate linear regression. Following MAPE errors are achieved for them: 7,8 % for low unit prices group and 4,4 % for high unit prices group. So high accuracy of predictions could be achieved thanks to meeting several requirements arisen from market conditions and machine learning rules, as well as, applying the hybrid approach. Wishing to apply the model to predictions of flats' prices, the database should be created based on the latest secondary estate market transactions and it should comprise cases recorded in the limited area i.e. the quarter of the city, the small town. Then, based on the area of a flat, the number of rooms, the storey number, the technical state of a block of flats, state of its vicinity, the price of a flat can be predicted with stated above MAPE error. The successful application of the proposed model based on historical data, except for current appraisals, can be applied for quick flats' valuation (e.g. for tax, or bank purposes), as well as for the optimization of a developer's pricing strategy, what is a plan of the authors.

REFERENCES

1. R. Gaca, "Parametric and non-parametric statistical methods in the assessment of the effect of property attributes on prices", Real Estate Management And Valuation 2, 28, 9-15, 2018

2. R. Gaca, "Comparative appraisal models", Rzeczoznawca majątkowy 2/2018, 8-13, 2018

3. International valuation standards (IVS) 2017

4. Ustawa z dnia 21 sierpnia 1998 r. o gospodarce nieruchomościami (Act of 21 August 1998 on real estate management).

5. TEGOVA European Valuation Standards, EVS 2016

6. E. Jahanshiri, T. Buyong, A.R.M. Shariff, "A Review of Property Mass Valuation Models", Pertanika J. Sci. & Technol. 19, 23 – 30, 2011

7. S. Kokot, S. Gnat, Simulative verification of the possibility of using multiple regression models for real estate appraisal, Real Estate Management and Valuation, vol. 27, no. 3, pp. 109-123, 2019

8. M. Renigier-Biłozor, "Application of the rough set theory to mass appraisal of real estate on small markets", Acta Scientiarum Polonorum. Administatio Locorum 7/3, 35-51, 2008

9. Rozporządzenie Rady Ministrów z dnia 21 września 2004 r. w sprawie wyceny nieruchomości i sporządzania operatu szacunkowego (Ordinance of the Council of Ministers of 21 September 2004 on the valuation of real estate and the preparation of an appraisal report).

10. M. Podwórna, W. Mironowicz, "Praktyczne aspekty oceny stopnia zużycia obiektów budowlanych, dla potrzeb wyceny nieruchomości", Rzeczoznawca Majątkowy Nr 1/2017 (93) s. 8-13, 2017

11. M.T. Hagan, H.B. Demuth, M.H. Beale, O. De Jesús, Neural Network Design; Martin Hagan: Lexington, KY, USA, 2014.

12. M. Kaftanowicz, M. Krzemiński, Multiple-criteria Analysis of Plasterboard Systems, Procedia Engineering 111:364-370, 2015, doi: 10.1016/j.proeng.2015.07.102

13. H. Anysz, A. Zbiciak, N. Ibadov, The Influence of Input Data Standardization Method on Prediction Accuracy of Artificial Neural Networks, Procedia Engineering 153, 2016, doi: 10.1016/j.proeng.2016.08.08

14. I. Foryś, R. Gaca, Theoretical and practical aspects of qualitative variable descriptions of residential property valuation multiple regression models. In: Proceedings, Foundation of the Cracow University of Economics, Cracow, 978-83-65173-48-5, pp 36–44, 2016

15. W.J. McCluskey, R.A. Borst, The Theory and Practice of Comparable Selection in Real Estate Valuation. In: d'Amato M., Kauko T. (eds) Advances in Automated Valuation Modeling. Studies in Systems, Decision and Control, vol 86. Springer, 2017

16. M. Ciuna, M. De Ruggiero, B. Manganelli, F. Salvo, M. Simonotti, Automated Valuation Methods in Atypical Real Estate Markets Using the Mono-parametric Approach. In: Gervasi O. et al. (eds) Computational Science and Its Applications – ICCSA 2017. ICCSA 2017. Lecture Notes in Computer Science, vol 10406. Springer, 2017, doi: 10.1007/978-3-319-62398-6

17. F. Tajani, P. Morano, M. Locurcio, N. D'Addabbo, Property valuations in times of crisis: artificial neural networks and evolutionary algorithms in comparison. In: Gervasi, O., Murgante, B., Misra, S., Gavrilova, M.L., Rocha, A.M.A.C., Torre, C., Taniar, D., Apduhan, B.O. (eds.) ICCSA 2015. LNCS, vol. 9157, pp. 194–209. Springer, 2015. doi: 10.1007/978-3-319-21470-2

18. M. Rogalska, P. Wolski, Prognozowanie cen 1m2 mieszkania na rynku pierwotnym w Warszawie metodą uogólnionych modeli addytywnych, Logistyka 6/2014, pp 9101-9110, 2014

19. M. Rogalska, Wycena nieruchomości z wykorzystaniem metody wieloczynnikowych modeli statystycznych, (chapter in:) Inżynieria Przedsięwzięć Budowlanych. Problemy Modele Metody, (edit. Sobotka A., Radziszewska-Zielina E.), KILiW PAN, 2018

20. N. Ibadov, H. Anysz, Ustalenie stopnia zużycia technicznego obiektów budowlanych w zarządzaniu nieruchomościami, Technika Transportu Szynowego. Koleje, Tramwaje, Metro 10/2013, pp. 49-56, 2013

21. B. Nowogońska, Diagnoses in the Aging Process of Residential Buildings Constructed Using Traditional Technology, Bulidings, 2019, doi: 10.3390/buildings9050126

22. B. Kutera, H. Anysz, The methodology of technical due diligence report preparation for an office, residential and industrial buildings, MATEC Web of Conferences 86:07009, 2016, doi: 10.1051/matecconf/20168607009

23. M. Juszczyk, A. Leśniak, Modelling construction site cost index based on neural network ensembles. Symmetry 11(3), 2019

24. M. Juszczyk, A. Leśniak, K. Zima, ANN Based Approach for Estimation of Construction Costs of Sports Fields, Complexity (1), 2018, doi: 10.1155/2018/7952434

25. M. Rogalska, Wieloczynnikowe modele w prognozowaniu czasu procesów budowlanych, Lublin University of Technology, Lublin 2016, ISBN: 978-83-7947-186-7

26. H. Anysz, N. Ibadov, Neuro-fuzzy predictions of construction site completion dates, Technical Transactions 6/2017, Civil Engineering, pp. 51-58, 2017, doi: 10.4467/2353737XCT.17.086.6562

27. H. Anysz, Wykorzystanie sztucznych sieci neuronowych do oceny możliwości wystąpienia opóźnień w realizacji kontraktów budowlanych, Oficyna Wydawnicza PW, Warsaw, 2017

28. L. Zajączkowski, Z. Waszczyszyn, A. Tomana, M. Jakubek, Sztuczne sieci neuronowe w wycenie działek budowlanych. (chapter in:) Konferencja Naukowa Komitetu Inżynierii Lądowej i Wodnej PAN i Komitetu Nauki PZITB (49:2003; Warszawa-Krynica), Problemy naukowo-badawcze budownictwa, ISBN 837207433X, Oficyna Wydawnicza PW, Warsaw, 2003

29. J. Morajda, Wykorzystanie perceptronowych sieci neuronowych w zagadnieniu wyceny nieruchomości. Zeszyty Naukowe Małopolskiej Wyższej Szkoły Ekonomicznej w Tarnowie, (7), 101-108, 2005

30. T. Jasiński, A. Bochenek, Prognozowanie cen nieruchomości lokalowych za pomocą sztucznych sieci neuronowych. Studia i Prace WNEiZ US, 45, 317-328, 2016, doi: 1018276/sip.2016.45/1-25

LIST OF FIGURES AND TABLES:

PROGNOZOWANIE WARTOŚCI RYNKOWEJ JEDNORODNYCH NIERUCHOMOŚCI HYBRYDOWYM MODELEM Z WYKORZYSTANIEM SZTUCZNYCH SIECI NEURONOWYCH

Słowa kluczowe: wycena nieruchomości, sztuczne sieci neuronowe (ANN), wielowarstwowy perceptron, podejście porównawcze, uczenie maszynowe, modele hybrydowe

STRESZCZENIE:

Wycena nieruchomości jest złożonym procesem. Rzeczoznawca majątkowy musi być biegły zarówno w naukach ekonomicznych, prawnych, jak i technicznych. W praktyce często zdarzają się przypadki, w których konieczne jest poznanie zakresu wartości nieruchomości w krótkim czasie. Zautomatyzowane modele wyceny (AVM) są kwestionowane przez praktyków, ale nie oznacza to, że nie należy szukać nowych metod wyceny, innych niż te określone w Rozporządzeniu Rady Ministrów z dnia 21 września 2004 r. w sprawie wyceny nieruchomości i sporządzania operatu szacunkowego. Do określenia wartości rynkowej nieruchomości zdefiniowanej w Ustawie z dnia 21 sierpnia 1997 r o gospodarce nieruchomościami, jako „szacunkowa kwota, jaką w dniu wyceny można uzyskać za nieruchomość w transakcji sprzedaży zawieranej na warunkach rynkowych pomiędzy kupującym a sprzedającym, którzy mają stanowczy zamiar zawarcia umowy, działają z rozeznaniem i postępują rozważnie oraz nie znajdują się w sytuacji przymusowej", najczęściej stosowaną metodą wyceny jest podejście porównawcze polegające na szacowaniu wartości na podstawie ostatnich danych sprzedaży innych podobnych nieruchomości na rynku lokalnym. Takie podejście wymaga aktywnego, rozwiniętego oraz w miarę stabilnego rynku. Rzeczoznawca majątkowy analizuje ceny transakcyjne nieruchomości, które w wystarczającym stopniu są podobne do nieruchomości wycenianej. Analiza atrybutów nieruchomości polega na badaniu nieruchomości pod względem trwałych cech, które mają znaczący wpływ na wartość, w szczególności lokalizację obiektu, jego powierzchnię, położenie w budynku, stan techniczny.

W pracy przeanalizowano próbkę 222 nieruchomości lokalowych, które były przedmiotem obrotu na wrocławskim rynku wtórnym. Lokalny rynek nieruchomości przyjęto jako nieruchomości lokalowe o powierzchni użytkowej z przedziału od 15 do 95 m2, w budynkach o stanie dobry lub średnim, z obrębu Grabiszyn dzielnicy Fabryczna miasta Wrocław. W pracy przyjęto dwuletni okres analizy, ze względu na w miarę stabilny rynek w okresie 2013-2014 nie uwzględniono czynnika czasu – przyjęto zerowy trend czasowy dla transakcji wolnorynkowych.

W oparciu o dane uzyskane z katastru miejskiego oraz badań własnych zbudowano bazę danych zawierające następujące informacje o każdym z lokali:

- cena transakcyjna w PLN,

- powierzchnia lokalu w m2

- liczba pokoi,

- piętro, na którym znajduje się lokal

- stan techniczny budynku (średni lub dobry),

- stan otoczenia budynku (otoczenie przeciętne lub komfortowe)

Na podstawie powyższych danych poszukiwano (programem Statistica 13.1) najlepiej prognozującej wartość lokalu sztucznej sieci neuronowej (warstwowej), w której na wyjściu była cena lokalu. Uzyskany średni błąd względny MAPE (ang. mean absolute percentage error) równy15,5 % wykluczał praktyczne zastosowania takich prognoz – był zbyt duży. Na podstawie wykonanego histogramu podzielono więc lokale z bazy danych na 3 podzbiory. Wyodrębniono podzbiór mieszkań L o niskich cenach 1 m2 (do 4000 PLN) o liczności 37, podzbiór M o cenach jednostkowych z zakresu 4000-6000 PLN/m2 (142 rekordy w bazie danych) oraz podzbiór H mieszkań o cenach 1 m2 powyżej 6000 PLN o liczności 43. Zbudowano sztuczną sieć neuronową ANN1, która bezbłędnie klasyfikowała mieszkania do właściwego podzbioru na podstawie w/w danych wejściowych. Ta 100 % trafna klasyfikacja stanowiła podstawę do rozbudowy modelu o drugi moduł prognozujący cenę mieszkań. Tylko dla podzbioru M – ze względu na wystarczająco dużą jego liczność można było ponownie zastosować sztuczną sieć neuronową jako narzędzie predykcji. Znaleziono 3 takie sieci, dla których MAPE nie przekraczał 6,5 %. Dla podzbiorów L i H zastosowano liniową regresje wieloraką (odrębnie dla każdego podzbioru). Uzyskano następujące błędy prognoz MAPE: 7,8 % dla podzbioru L oraz 4,4 % dla podzbioru H. Średni błąd MAPE dla całego modelu zważony liczbą przypadków w każdym podzbiorze wynosi 6,3 %, co jest wartością bardzo dobrą i wiarygodną. Mieszkania o identycznych cechach (zastosowanych w modelu) różnią się rodzajem wykończenia, wyposażenia, stanem prawnym. Tych cech proponowany model nie uwzględnia, więc uzyskanie zbyt niskich wartości MAPE świadczyłoby o przeuczeniu modelu lub innych błędach metodologicznych.

Stworzony dwustopniowy model prognozowania cen mieszkań może być stosowany do wstępnej – zgrubnej – wyceny wartości lokali z w/w błędem. Precyzyjna wycena, zarówno od strony merytorycznej, jak i formalnej, musi być poprzedzona oględzinami lokalu przez rzeczoznawcę i jego szczegółową, wielostronną oceną. Nie mniej, zaproponowany model wyceny – przede wszystkim ze względu jej szybkość i nieduży błąd – może znaleźć uznanie wśród inwestorów którzy coraz częściej stosują formalne metody ilościowe do konstruowania portfeli aktywów, które optymalizują oczekiwane ryzyko i zwrot z inwestycji. W artykule pokazano, że możliwe jest wykorzystanie narzędzi uczenia maszynowego do prognozowania cen jednorodnych nieruchomości, jednak by możliwe było szersze zastosowanie takiego narzędzia niezbędne jest posługiwanie się wystarczająco dużą baza danych dotyczącą rzeczywistych transakcji. W przypadku nieruchomości bardzo duże znaczenie ma jej lokalizacja. W analizowanym modelu ten problem ominięto analizując transakcje dotyczące lokali z części jednej z dzielnic miasta. Stosując podobne modele dla większej liczby lokalizacji, należy jednak brać pod uwagę także czynnik lokalizacji.