Reports on Geodesy and Geoinformatics, 2018, Vol. 106, pp. 25–34

DOI: 10.2478/rgg-2018-0011
Received: 27 August 2018 / Accepted: 5 December 2018
Published online: 10 December 2018

ORIGINAL ARTICLE

The use of urban indicators in forecasting a real estate value with the use of deep neural network

Anna Bazan-Krzywoszańska[1*] and Michał Bereta[2]

[1]Faculty of Civil Engineering, Architecture and Environmental Engineering, University of Zielona Góra, 1 Prof. Z. Szafrana street, 65-516, Zielona Góra, Poland
[2]Department of Computational Intelligence, Faculty of Physics, Mathematics and Computer Science, Cracow University of Technology, 24 Warszawska street, 31-155, Cracow, Poland

[*]A.Bazan@aiu.uz.zgora.pl

Abstract

Records of municipal planning documents directly affect the land use. In this way, the market price of the land is also shaped. Awareness of the economic and social consequences of adapting specific solutions is the primary argument that should condition the local policy in terms of spatial planning. The research results indicate that the network trained with attributes which do not describe a property value by its price was able to estimate it with acceptable and satisfactory results. The possibility to use artificial multilayer networks in spatial policy decision-making seems well founded. The research results show the relevance of the assumption that using them for modeling can be helpful in selecting the most advantageous variant of planning arrangements in a local law document which determines the land use and development, therefore impacts its value.

Key words: planning analysis, local zoning plan, local policy, deep learning, deep neural networks, machine learning, artificial intelligence

1 Introduction

The land designation made on the basis of a planning document directly affects the land use. In this way, the market price of the real estate is also shaped. Then, it should be assumed that a municipal spatial policy has a direct impact on the economic value of a neighborhood (Cymerman et al., 1999; Laskowska, 2008). Adoption of a local plan, resulting in a change of land designation, has an impact on a real estate market value, causing it to rise or fall. Thus, when implementing a planning document, which has the indicated effects, it is reasonable to assess the impact of its findings on the space, but not only in the context of the documents required by law, i.e. assessment of the impact on environment and assessment of the financial impact of the plan but also in the context of forecasting the influence of correlation of adopted indicators and parameters on the value of real estate covered by the plan. According to the intention of the legislature, the analysis of the planning document, not only in terms of designation (function) of a real estate, mentioned in the plan, but also taking into account the remained indicators gives knowledge about the actual effect of seemingly insignificant urban records. These actions should precede completion of the planning procedure, related to the adoption and implementation of the local law, that is the local zoning plan.

In the current state of knowledge, to determine the changes in a real estate value, regarding the intended land use specified in the local plan, the value of the property before and after the change is examined. The price is determined mostly on the basis of average transaction prices for properties of similar intended use (Łaguna et al., 2004; Cymerman, 2008).

The legislature does not require units responsible for implementing a municipal spatial policy to determine the economic effects of adopting specific indicators in the plan. This means that space management is virtually out of control for the indicated aspect, as the forecasts of financial effects and impact on environment made at the stage of completion of the plan, concern only the effects of its enactment, in search for taxes, that an investor shall be charged by when implementing the findings of the plan. These actions are carried out in accordance with the art. 36 par. 1 and 4 of the act on spatial planning and development (Act, 2017). They apply to cases in which the real estate value is increased or decreased, as a result of the adoption of planning records. If the real estate value increases due to the adoption of the plan, and the owner disposes this property within 5 years from the date of the plan's enactment, a commune administrator or a mayor shall charge the owner a one-time fee, determined in the plan. The so-called planning fee is determined as a percentage of the real estate value. Its height, according to the will of the legislator, may not be greater than 30% increase in the real estate value. In the case the real estate value decreases due to the adoption of the plan, the owner may apply for compensation for the decline in the real estate value.

Given the above, it must be stated that municipal spatial policy is shaped, by local laws, without consideration of the real effects of adopting planning records and their impact on the value of properties covered by the findings of the local plan. At the stage of implementing the plan and determining urban indicators, nobody controls the effects of the planned arrangements, e.g. using a comparative analysis of several concepts with different values of individual indicators assigned.

The aim of the research presented here is to analyze the possible use of findings of a planning document and conditions of land covered by a local plan, within the factors that may affect the real estate value, in order to build predictive models of real estate prices. Deep neural networks were used as predictive models and a linear regression model was taken for comparison. The use of such models for spatial planning issues could bring undoubted benefits in terms of knowledge of a real estate value, which is influenced by findings of a planning document, as well as abilities to program and somehow control provisions of a local plan at the stage of implementation of the document.

Real estate price appraisal using neural networks is one of the methods adopted by researchers to estimate land value. Application of neural network in relation to the traditional method is considered a faster way of obtaining results (Leśniak and Juszczyk, 2018). Currently, estimates and appraisals for real estate valuations are based on the guidelines set out within the regulation on the property valuation and appraisal reports (Regulation, 2004).

The work presents the application of neural network in estimating a real estate value not only by using features, that enable to describe them in space but also by urban indicators, which are part of a planning document. As a result of the research, neural networks show their ability to be trained to the extent that allows assuming that planning document findings, not only in terms of land designation, directly impact on real estate value, and the network itself can be a useful predicting tool. On the basis of the adopted data (including urban indicators), networks show the ability to be trained effectively and thus predict the value of a property. General diagram of actions in the adopted model to support urban planning is shown on Fig. 1.

The following methods are used in forecasting changes in a real estate value: statistical, using decision trees and neural networks, taking into account the characteristics of the local market (Laskowska, 2008). Approach using a comprehensive system of prediction, using Geographic Information System (GIS) and models of artificial neural networks to the real estate valuation was proposed, inter alia, by García et al. (2008). To improve the efficiency of a real estate valuation, researchers combine the real estate appraisal system based on GIS and neural network trained with backpropagation algorithm, in order to construct a model for appraising the price. This method, due to the search function, spatial data and attribute edition, improves the efficiency and accuracy of the valuation so that the new method of real estate is provided by the system (Liu et al., 2011).

Neural networks are considered by the current science as a popular multidimensional analytical tool, inter alia, on the need for a rational valuation of real estate (Minli and Yueran, 2010). Scientists are also trying to apply analyses of the Neural Network predictive capacity in conjunction with Multiple Regression Analysis (MRA) (Xu et al., 2017; Wang et al., 2018). The integration of these systems has become the subject of research, also in terms of efficient real estate valuation. This type of integration undoubtedly reduces restrictions, using the capabilities of systems that give more options than a single intelligent system, if combined. For this reason, scientists propose the use of hybrid systems for efficient and more precise real estate valuation than a traditional one (Taffese, 2007). Subjective decisions in the real estate valuation are also corrected with coefficients within a method called Quantitative Comparative Approach. The principle is the assumption that the price per unit area of real estate is the average price per unit area of the particular set of housing supply and demand multiplied by the product of several dimensionless adjustment coefficients of factors (Yeh and Hsu, 2018).

Research carried out by Jasiński and Bochenek (2016) modeling the sales price of real estate was developed using GIS data, archived sales prices with the use of additional information obtained, inter alia, in the process of analysis of land and mortgage registers. The literature review showed models based on a relatively small number of explanatory variables, as well as those built on a large number of data could both work correctly (Jasiński and Bochenek, 2016). Independently, neural networks are used in prediction models of decision-making about interior design, taking into account the economic, social and environmental effects (Bazan-Krzywoszańska et al., 2019).

The possible use of neural networks in spatial development, presented in the literature, relate only to the valuation and appraisal of real estate value. The model presented in the work uses the characteristics of the space development and equipment, as well as parameters and urban indicators, established in the local plan project, which may impact the functioning and the possible use of real estate, and thus on its value. Additionally, it must be noted that this model, using the deep neural network, was trained without the attributes that describe the real estate value, such as average market price.

2 Legal status

In accordance with the art. 1 par. 3 the Act of 27 March 2003 on spatial planning and development (Act, 2017) the authority responsible for the implementation of municipal spatial policy (Municipal Council through Commune Administrator, or Mayor), and thus for the implementation of local zoning plans, while specifying land use and its development, is required to consider the public interest and private interests. Therefore it ought to protect the existing land use as well as to change it considering economic, environmental and social analyzes. The requirement of efficient space development, considering the economic value of space, in accordance with the art. 1 par. 4 of the Act is to be carried out by appropriate shaping of spatial structures.

Independently, according to the art. 15 par. 1 subpar. 3 of the Act, the plan's draft justification shall determine its impact on the public finances, including the municipal budget, pointing to the art. 15 par. 2 subpar. 12, as the basis of the

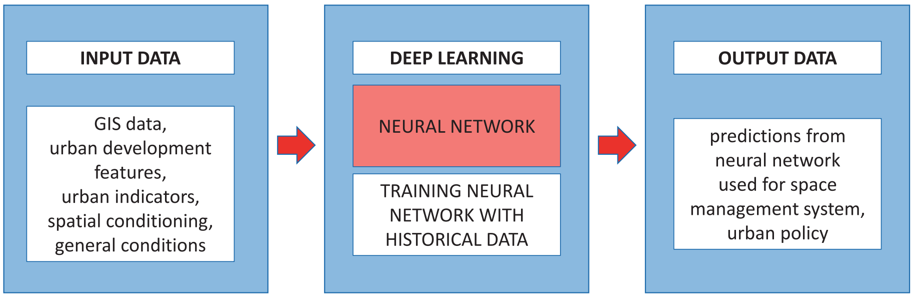

Figure 1. The analysis of the real estate value on the basis of the conditions arising from the findings of the planning document

percentage which decides on the fee resulting from the art. 36 par. 4 of the Act – that is resulting from the increase in the value of the property.

Based on the art. 37 par. 12 of the Act, to the rules for appraising the real estate value and the financial consequences of the adoption of the local plan shall apply the art. 151 par. 1 and the art. 152, in conjunction with the art. 154 of the Real Estate Management Act of 21 August 1997 (Act, 2018). Methods and techniques, inter alia, to appraise the real estate value, considering all the necessary and available data about the real estate, referred to in art. 155 of the Act, are determined based on the regulation of the Council of Ministers of 21 September 2004 on the real estate valuation and appraisal (Regulation, 2004).

Given the above, it should be noted that none of the listed elements of financial analysis, referred to in the Act on spatial planning and development in conjunction with the specific provisions, refers to the stage of preparing the plan. This is a stage of analyzing results and impacts of specific solutions e.g. variants related not only to the land intended use or function but also to other circumstances which may affect the real estate value if a certain variant of the plan's records is adopted. The model presented in the work shows the possibility of using the deep neural network in obtaining knowledge on real estate value, which is effected by the planning document's records.

3 Artificial neural networks

Multilayer artificial neural networks are currently among the most effective and widely used calculation methods of artificial intelligence (Leśniak and Juszczyk, 2018; Ertuğrul, 2018). Their popularity has increased in recent years, mainly as a result of the development of more efficient network training algorithms with many hidden layers, called *deep learning* (Goodfellow et al., 2016; Hofmann and Klinkenberg, 2013).

The primary component of the neural network is a neuron. Mathematically, it is a weighted combiner, which outputs a nonlinear function called activation function. Fig. 2 shows the functional diagram of a neuron where the vector $\mathbf{x}$ is a vector of N input signals, and vector $\mathbf{w}$ is a vector of weights, that can strengthen or weaken the signals passing by. Weight $\mathbf{w}_0$ is associated with so-called bias signal, with a fixed value of 1.

Currently, ReLU (rectified linear unit) is an often used type of neuron. This is a type of neuron commonly used in deep neural networks. It was also used in the in this work in calculation research. For ReLU neuron activation function f is defined as

$$f(\mathbf{x}) = \max(0, \mathbf{x})$$

This function is shown on Fig. 3. In the most common type of neural networks – multilayer perceptrons, neurons are combined in layers. Only the first layer receives the input signals, that is, the attribute values on the basis of which the network has to make decisions in a given problem. Connections between

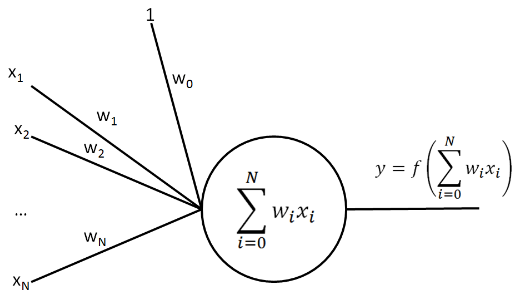

Figure 2. Functional diagram of a neuron

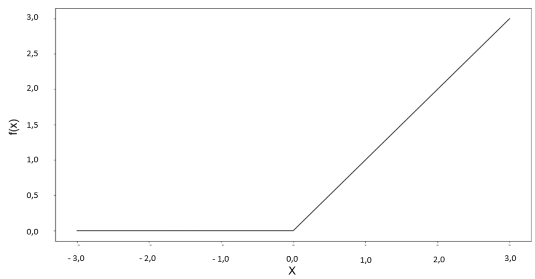

Figure 3. Activation function of ReLU neuron

neurons only occur between successive layers which are fully connected. Output signals of the last layer neurons are the response of the whole network. The number of neurons in this layer depends on the addressed problem (Amakdouf et al., 2018; Goodfellow et al., 2016). For example, in the issue considered in this paper, the price of real estate is estimated. Therefore one neuron in the output layer is needed, and its output signal, calculated on the basis of the input signals of the whole network being processed by successive hidden layers, provides the estimated value. This is, therefore, a linear neuron, not a ReLU. The hidden layers can contain any number of neurons. Their number may be subject to optimization depending on the issue (Goodfellow et al., 2016; Zhang et al., 2003). The sample diagram of a network with three hidden layers is presented on Fig. 4. The neural network as a whole can be interpreted as a system that performs a multiple variables function

$$y = f(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ is a vector of attribute values (input signals), on the basis of which the decision is taken; y is the output signal of the network, in this work it is the expected price of the real estate. The properties of artificial neural networks allow their practical applications to solve such issues as: recognition and classification of patterns, approximation, prediction and many more (Mrówczyńska, 2015).

Table 1. Urban indicators and data used for the model, including those which affect the real estate investment attractiveness

Endogenous factors		Exogenous factors
• address of the property, • type of real estate, • function according to registry of property prices and values, • area, number of floor, number of storeys in the building, • storey of the property, • building height, • state of the property, • parking lots, • availability of technical infrastructure, • primary and secondary intended use in the local zoning plan,	• alternative use in the zoning plan, • a form of conservation protection, • maximum size, • minimum biologically active size, • the value of the property tax resulting from the art. 36 of the Act (2017), • the number of parking spaces within the real estate, • a form of possession, • wooded areas, • the form of greenery protection, • visual appraisal, • planning fee;	• district safety, • distance from a bus stop and basic services, • forms of transport, • the condition of the road, • the availability of parking spaces, • insolation, • additional restrictions resulting from the provisions of separate regulations, • nuisance;

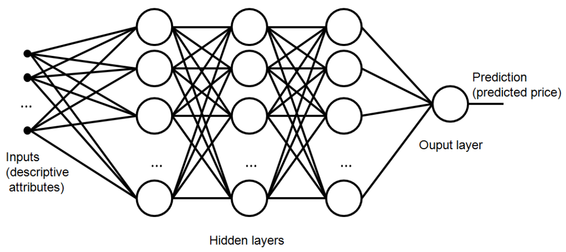

Figure 4. Neural network with three hidden layers and one output neuron for price prediction

3.1 The training process of a neural network

For fixed network architecture (the number of layers and the number of neurons in each layer), non-linear function performed by the network depends on the weights of connections between neurons. The process of selecting values of these weights is called network training. It takes place on the basis of the available historical data representing the known cases of the addressed problem (called a training set). In the issue of estimating real estate prices, a training set contains historical transactions with a known sale price of the estates described with a certain set of attributes, such as size or location. The training process is iterative. The network is presented with successive examples as values of attributes provided to the inputs of neurons of the first layer, then the answers of subsequent layers are calculated until the last layer, where the output signal is the network's response to the solved problem. This response is compared with the actual, known for examples from a training set. If the network's answer is not satisfactory (mistake is too big) a modification of the network's weights take place, in such a way that at the next presentation of the example, the error is smaller.

The majority of the neural network training algorithms belong to gradient methods of training (Bataineh and Marler, 2017). However, their detailed explanation exceeds the capabilities of this work. Nevertheless, it is worth mentioning several important aspects of the network training process. The training process is iterative, so each training example is presented to the network repeatedly. A training set should contain as many examples as possible. The aim is to select such a set of weights to make the network response with the smallest possible error for test examples, that is for cases not presented while training (Goodfellow et al., 2016). In the example of predicting real estate prices, those may be new potential sales of some real estate, where the final prices are not known yet. Knowing, however, the values of attributes that describe the property, one can calculate the response of the network which is an estimate of the price.

The network training process is usually accompanied by assessing the quality of the resulting model. All available historical examples are divided into training data and test data (sometimes called validation data). For both, the correct (desired) response is known. However, only training samples affect the process of modification of the network weights. Test samples are only used to check the network response quality. The process of assessing the model quality can be more complicated. Cross-validation, as described in the further part of this work, is the example. To obtain the best model for a given problem, there is often a need to repeatedly train networks with different architectures, activation functions, or values of training algorithms parameters.

3.2 Selection of input attributes

Selection of the appropriate input signals (that is, attributes describing each historical case) is crucial for the predictive ability of networks. During their preparation, two aspects are important:

- input signals should contain the appropriate information regarding the addressed problem; In many cases, their selection should be supervised by an expert;
- each neuron requires that signals coming to it have certain values, all attributes must be numeric; If in the issue, certain attributes are nominal, they must be properly coded numerically. The manner it was performed in the considered issue is described in the further part of this work.

In the issue of forecasting real estate prices, in order to create the model, data and factors were selected which had an impact on the investment attractiveness of real estate, with regard to the conditions and indicators of the local plan that impact the value and thus the attractiveness of real estate (Szewrański et al., 2017; Mrówczyńska et al., 2018; Juszczyk et al., 2018). Conditions, indicators and arrangements of the local plan in the process of predicting real estate prices can be divided into groups (Jasiński and Bochenek, 2016) i.e. endogenous and exogenous (Table 1).

The data used for modeling were obtained from 163 sale and purchase transactions for 2016 and 2017, concluded within the city center of Zielona Gora (Lubuskie voivodeship, Poland), originating from the registry of prices and the value of real estate held by the Mayor of Zielona Gora. Estimated value (de-

Table 2. Numerical attributes used in the research

No.	Attribute name	Min	Max	Average	Deviation	Miss. val.*
1	plot area [m^2]	51	10717	1069.82	1601.71	0
2	building premise area [m^2]	8.91	6822.9	157.19	691.47	15
3	max. developed area [m^2]	0.3	1	0.67	0.26	0
4	min. biologically active area [m^2]	0	5	0.22	0.67	1
5	tax value of the art. 36 [%]	25	30	30	0.01	23
6	planning fee [PLN]	8594.95	414255	91567.72	65867.32	0
7	adopted price per 1 m2 [PLN]	1539.74	5122	3817.62	708.78	0

* Examples with missing values

Table 3. Nominal and binary attributes used in the research

No.	Attribute name	Num. of values	Miss. values
1	address	44	0
2	function registry	15	0
3	the storey of the property	7	10
4	property with land share	2	0
5	valid Local Zoning Plan	2	0
6	Zoning Plan primary function	11	12
7	function registry for the function of Local Zoning Plan	2	21
8	Zoning Plan secondary function	2	2
9	Zoning Plan alternative function	2	0
10	a form of conservator's protection	4	0
11	conservator's protection zone	4	0
12	allowable demolition	2	0
13	build-in height	4	0
14	access	4	0
15	the condition of the access road	3	0
16	access to green areas	2	0
17	availability of parking spaces	3	0
18	availability of basic services	2	0
19	availability of public transport	2	0
20	need to bear financial expenses of restoring	2	0
21	form of possession	2	0
22	form of functioning	2	0
23	technical condition of the building	3	0
24	wooded areas	2	0
25	protective greenery	2	2
26	insolation	3	0
27	extra restrictions	2	0
28	district safety	3	0
29	functionality	2	0
30	visual appraisal	3	0
31	noise	2	0
32	nuisance	2	0
33	additional functions according to plan	5	0
34	Zoning Plan secondary function	2	2

pendent variable) was the real estate value in PLN/m^2 .

The study uses data for 2016 and 2017, as older ones do not reflect current realities about property prices in Zielona Gora. In the selected test area and period Zielona Gora City Hall registered 163 transactions, meeting the location and real estate requirements, in the system. The data from the registry apply to prices of real estate along with the land participation which arises from notarial deeds of purchase/sale of real estate, which are the basic documents to build the official registry. The data collection was boosted by indicators and data which characterize a property in terms of its investment attractiveness, referred to in Table 1. In Table 2 numerical attributes used in the research are presented. Among them, there is the price for m^2, which is the target feature predicted by the proposed model. In Table 3 nominal and binary attributes are presented.

In the case of binary attributes they are encoded numerically as the values from the set {-1.1}, where "-1" is encoded `false` value , "1" is encoded `true` value.

The nominal attributes are encoded as a set of new binary attributes and encoded numerically the same as in the case of binary attributes. For example, if the nominal attribute `Visual appraisal` takes the possible values: `nice`, `ugly` and `with no special features`, for the purposes of the research they are changed to three binary attributes:

- `visual_appraisal_nice`,
- `visual_appraisal_ugly`,
- `visual_appraisal_no_special_features`.

When the value of the original nominal attribute `Visual` is `ugly`, corresponding numeric attributes have the following values:

$$\begin{aligned} \texttt{visual\_appraisal\_nice} &= -1, \\ \texttt{visual\_appraisal\_ugly} &= 1, \\ \texttt{visual\_appraisal\_no\_special\_features} &= -1. \end{aligned}$$

By adopting this procedure for encoding all the attributes used for the model are shown as numeric.

Only those attributes were used that do not have direct information about the price in PLN/m^2 (new polish złoty/m^2). The aim was to examine whether deep neural networks are able, on the basis of these data, to predict the real estate price in PLN/m^2. When the attributes were imported in this way, 163 examples and 41 original attributes (numerical, nominal and binary) describing individual real estates were obtained. The original nominal and binary attributes were then encoded in the previously described manner. Eventually, 161 numeric attributes were received, many of them with the values from the set {1,-1}. Missing values have been replaced with a value of 0.

3.3 Applied models

The study compared the following models:

i. a linear model, included in this study as a reference method; the linear model was the M5 model proposed by Quinlan (1992). M5 constructs tree-based piecewise linear models which combine a standard decision tree with the linear regression functions at the nodes.
ii. neural network with one hidden layer, 100 ReLU neurons, trained by 10000 epochs, ADADELTA training algorithm;
iii. neural network with two hidden layers, 100 and 70 ReLU neurons, trained by 10000 epochs, ADADELTA training algorithm;
iv. neural network with three hidden layers, 100, 70 and 50 ReLU neurons, trained by 10000 epochs, ADADELTA training algorithm;

In the above statement, the linear model was used as a kind of reference model. Models of neural networks as nonlinear models, will be compared with the linear model in order to check whether the application of neural networks is meaningful, that

Table 4. Results for split validation; MAE – Mean Absolute Error; MRE – Mean Relative Error

Model	Training		Testing	
	MAE [PLN/m^2]	MRE [%]	MAE [PLN/m^2]	MRE [%]
Linear regression	0.00 ± 0.00	0.00± 0.00	254.50 ± 284.57	7.22 ± 9.20
Neural Network (100)	3.30 ± 5.08	0.09 ± 0.15	160.63 ± 249.72	4.90 ± 10.01
Neural Network (100-70)	1.96 ± 3.02	0.05 ± 0.08	123.16 ± 215.32	3.95 ± 9.60
Neural Network (100-70-50)	0.96 ± 1.83	0.03 ± 0.08	83.49 ± 209.38	2.20 ± 5.23

is, whether it brings any benefit compared with the known and much simpler model.

Evaluation of the model was based on two values:

i. Mean absolute error (MAE) in PLN/m^2. The average absolute deviation of the prediction from the actual value, given as

$$MAE = \frac{1}{M}\sum_{i=1}^{M}\left|y_{pred} - y_{real}\right|, \tag{2}$$

where y_{pred} is the price predicted by the model, y_{real} is the true value of the price, M is the number of transactions in the training/testing set.

ii. Mean relative error (MRE) in %. The average relative error is the average of the absolute deviation of the prediction from the actual value divided by actual value, given as

$$MRE = \frac{1}{M}\sum_{i=1}^{M}\frac{\left|y_{pred} - y_{real}\right|}{y_{real}} \cdot 100\%. \tag{3}$$

Modeling was carried out using RapidMiner Studio 7.6 (Hofmann and Klinkenberg, 2013).

4 Research results

The numerical experiments, carried out for the purposes of research, were performed in two stages. In the first stage, a division of all available examples of historical training data and test data was made in the ratio of 60%/40%. We used the 60/40 split ratio instead of more common 70/30 or 80/20, as we wanted to use more examples for testing to make the evaluation more reliable. We are aware, that with such a small number of available historical transactions, there is a problem with constructing big enough training and testing sets. In our opinion having a smaller number of testing examples makes split validation less reliable in this case. In the second stage, the tests were performed with the cross-validation procedure, widely used in machine training. It is intended to provide a better evaluation of the average performance of the models.

4.1 Results for split validation

Table 4 shows the average results obtained for the training and testing data (along with the standard deviation specified as ±), for the considered models with available data randomly divided in the ratio 60%/40% (training/test). The applied procedure is characterized on Fig. 5.

Larger error values for test data in comparison with the results of the training data is a typical result in machine learning. Modern methods of training neural networks have built-in mechanisms of so-called regularization to prevent the adverse over-fitting of the network, i.e. undesired reduction of training error at the expense of increased test error. It is worth recalling that the test data results enable to assess the utility of a model in the new cases. Virtually zero train error for a

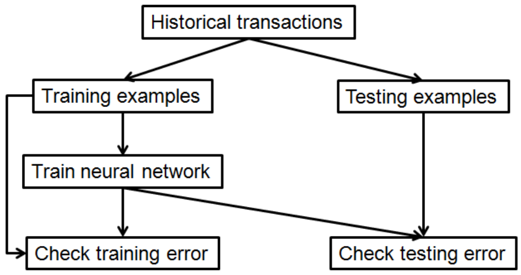

Figure 5. Illustration of the split validation procedure

linear regression model is nothing positive in the face of its much higher error in test data. On the other hand, neural networks have a higher error in the training data, but at the same time, they can better generalize knowledge in a training set as their test error is smaller. This is an important argument for the application of neural networks. The used linear regression model, M5, constructs tree-based piecewise linear models which combine a standard decision tree with the linear regression functions at the nodes. In this task, M5 as a tree-based model is exposed to overfitting during the training stage, even more than neural networks, as our results show. The modern regularization algorithms for neural networks protect the neural model from overfitting more effectively in case of our limited data.

The best result, measured by a relative test error, was achieved by the largest neural network (three hidden layers). This error was only avg. 2.20%, which is about 84 PLN difference between the actual price and expected price. These results confirm the possibility of using the deep neural network to predicting real estate prices on the basis of a data set composed of urban indicators and arrangements in a planning document, which is a local zoning plan. The research confirmed that neural networks could be a tool used at the stage of implementation of the local plan's draft, before the completion of the work on this document. Test error of a model decreases, as the network is growing. This means that the use of deep learning is more efficient compared to simpler networks.

In addition, Fig. 6 and Fig. 7 show histograms of the distribution of absolute errors for linear regression, and a network with three hidden layers. A number of test transactions (axis OY), for which estimating the price was an absolute given error (OX) can be read from it. One can observe that, although both models happened to make predictions with an error of 1000PLN/m ^{2}, the neural network provided far more cases with lower errors than the linear regression model. Similar conclusions can be drawn on the basis of the Fig. 8 and Fig. 9 where histograms of the differences between the expected and the actual price for the test data are presented. Therefore, the positive values on the axis OX relate to transactions, where the predicted price was overvalued in relation to the real price, while negative values relate to transactions where the model underestimated the price in relation to the actual one.

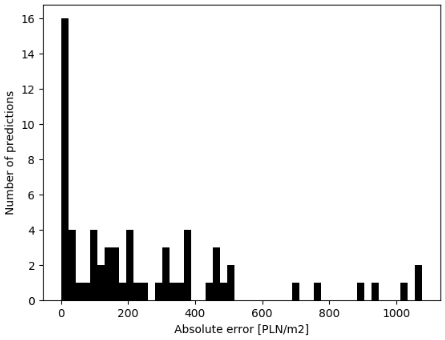

Figure 6. Histogram of absolute errors for test data for linear model

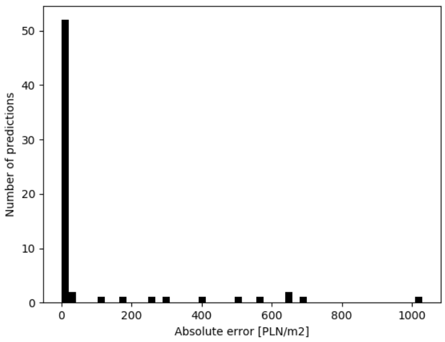

Figure 7. Histogram of absolute errors for test data for the neural network model with three hidden layers

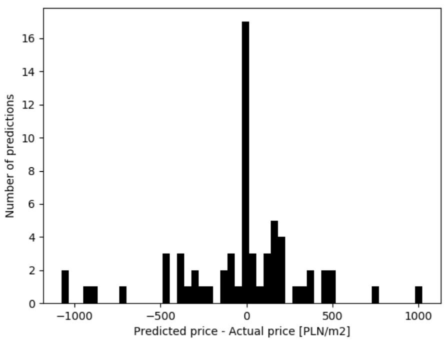

Figure 8. Histogram of differences between the predicted and actual prices for test data for the linear model

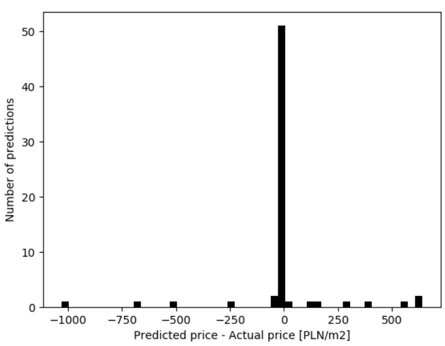

Figure 9. Histogram of differences between the predicted and actual prices for test data for the neural network model with three hidden layers

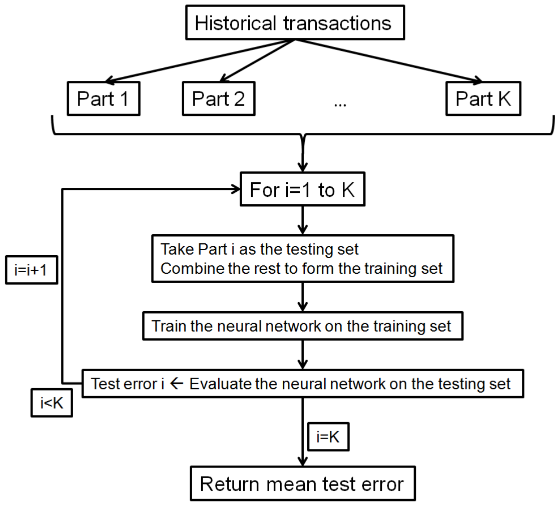

Figure 10. Procedure of cross-validation

Table 5. Results for 10-fold cross-validation; MAE – Mean Absolute Error; MRE – Mean Relative Error

Model	Testing errors	
	MAE [PLN/m^2]	MRE [%]
Linear regression	365.00 ± 143.48	12.94 ± 7.18
Neural Network (100)	141.59 ± 49.98	4.50 ± 3.03
Neural Network (100-70)	124.73 ± 109.06	4.08 ± 4.77
Neural Network (100-70-50)	125.75 ± 90.11	3.68 ± 3.24

4.2 Results for cross-validation

In the case of a relatively small data set which we are dealing with in this research, dividing it into two parts, as described in the previous point, makes the estimation very unstable. Also, the training process does not have a sufficient number of examples. Therefore, cross-validation is often used to get the average rating of the quality of the given classification or predictive model. In this method, the original set of all available historical data is split into K subsets. Then each of them is used as a test set, and the remaining K-1 parts constitute the training set. Analysis (training and testing) is, therefore made K times. K results are then averaged to get one result that is the estimate of the model performance. The application of cross-validation also allows verifying statistical significance of the observed differences in the operation of the individual methods. This method has been shown schematically on Fig. 10.

Table 5 shows the results of 10-times ($K = 10$) cross-validation for these models. The results confirm the previously observed relationship between the models, which is an advantage of neural networks on the linear model and the benefits of increasing the size of a network. However, due to the used procedure, this estimation of the model quality can be considered more reliable.

The results of cross-validation were the subject of statistical analysis by ANOVA tests and Student's t-distribution available in RapidMiner software. In simple terms, the ANOVA test null hypothesis implies that there are no statistically significant differences between the achievements of the various methods. The ANOVA test result in this case was the rejection of the null hypothesis, then the conclusion is, that there are statistically significant differences between the models. The null sources of these differences can be searched for performing the Student's t-distribution for each pair of models. Null hypoth-

Table 6. Relative importances of the first ten attributes selected based on the weights of the first two layers of the trained neural network with three hidden layers

Variable	Relative importance
function_registry_garage	1.000000
function_registry_residential_building	0.800775
function_registry_warehouses	0.480549
planning_fee	0.412057
storey_of_the_property_5	0.262469
function_registry_administration	0.233232
district_safety_1	0.215955
function_registry_commercial_service	0.199752
address_Lipowa	0.195060
insolation_1	0.154086

esis for a given pair of methods assumes that between these two methods there is no statistically significant difference in operation. The analysis at the significance level 0.05, showed significant differences between the linear model and any of the neural networks. However, it indicated no statistically significant differences for any of the networks pairs. This means that linear regression is worse from any neural network and this is a statistically significant difference. The test was not able to detect statistically significant differences between the networks. However, increasing the depth of the network tends to reduce the relative error.

The presented results show that the use of deep neural networks is an effective method of predicting real estate prices. The best considered model had the mean relative error of prediction 3.68% $\pm$3.24%, which translates into avg. 125 PLN error in predicting the price per m^2. It must be noted that none of the attributes describing the value, such as the average market price for similar properties, were used. The result can be considered as satisfactory and allows to look at the further use of networks in predicting other urban indicators such as investment potential of real estate.

4.3 The importance of the input attributes

Neural networks are referred to as ‘black box’ models. This means that providing the values of certain attributes to the input, and then receiving the response as signals of the output layer neurons, does not enable to directly interpret and follow the decision-making process. Specifically, the impact of various attributes on the final response cannot be directly indicated. Especially for a network with many hidden layers, the dependence of the output signal is a strongly non-linear function of the input attributes. However, there are some heuristic methods, trying to determine the significance of individual attributes assigned by a network. The simple idea is to recognize those connections between neurons which have a higher absolute weight value at the end of training, as more significant. The model used in RapidMiner is able to assess the appropriateness of individual input attributes, on the basis of weights of neurons in the first two layers. Table 6 presents a list of the ten most important attributes according to this method, along with the assigned numeric significance ratings (based on a trained network with three layers used in split-validation experiments). In many ways, this is an interesting list, as one can see the network takes into account information about the characteristics of the place, its location, intended use, but also additional features such as insolation, or safety. It is worth noting that the network does not make a decision on the basis of only those conditions. However, these are heuristically recognized as the most significant. There is no simple answer to the question of why some of them are top rated and not the others. This is the result of the network training when even thousands of weights of neural connections are selected. An interesting direction of future studies will be to compare the assessment of the significance of attributes identified by a neural network as the most important, and those identified by human experts.

5 Summary

The aim of the presented research was to analyze the possible use of a planning document findings and conditions of the land covered by a local plan, within the factors that may affect a real estate value, in order to build predictive models of real estate prices. Deep neural networks were used as predictive models and a linear regression model was taken for comparison. Neural networks with one, two and three hidden layers were used. Split validation was used to illustrate the model’s performance and cross- validation was used in order to estimate the performance more systematically and more reliably. Statistical test ANOVA and Student’s t-test were also used for a statistical comparison of results. The use of neural network models required adequate data preparation. The small number of historical transactions was the main problem. The vast part of the attributes was nominal, or binary and required appropriate encoding as numeric attributes. The data also missed some values, which were replaced by default ones. Despite difficult conditions, data preparation strategy allowed the proposed models to achieve promising results.

The research results suggest the relevance of the assumption that modeling, using deep neural network, can be a tool helpful in choosing the most advantageous variant of planning arrangements in a local law document, that determines the land use and development, thus has an impact on its value.

The presented computing studies indicate that the best model concerned in this work (neural network with three layers of hidden), which estimates the real estate value, had the mean relative error of prediction at 3.68% $\pm$ 3.24%, which translates into an avg. 125 PLN of error in predicting the price per m^2. The average real estate value within the adopted data was 3764 PLN/m^2.

The attitude towards a planning document presented in this work gives the opportunity to deal with the issue of planning records regarding their real effects. In this case, economic impact, related to the value of the property covered by the findings of the planning document. Urban indicators included in the study, confirm the assumption which demonstrates the existence of a significant relationship between the findings of the planning document and the value of the real estate covered by these records. This is due not only to the planning records that indicate function and use of a real estate. As is clear from the research, it is dependent on many other parameters that affect the use and equipment of a real estate (insolation, district where the property is located, technical infrastructure, access to roads, green system, services, biologically active surface rate, development area rate, development intensity, planning fee) that are in most cases not mentioned in the document required by law i.e. assessment of the financial impact of the plan adoption. This approach gives the opportunity to determine the hierarchy of factors affecting the value of the property.

The presented method allows to apply scenario analysis of the findings of the plan – take into account the impact of individual effects and indicators on the real estate economic value. In the present state of law, such an approach is not taken into account. The possible effects of the plan, with a big estimating error resulting mainly from the methods of forecasting the financial impact are considered after the plan completion.

Another effect of this type of actions is modeling space using a specific set of attributes. Visualization of the results of

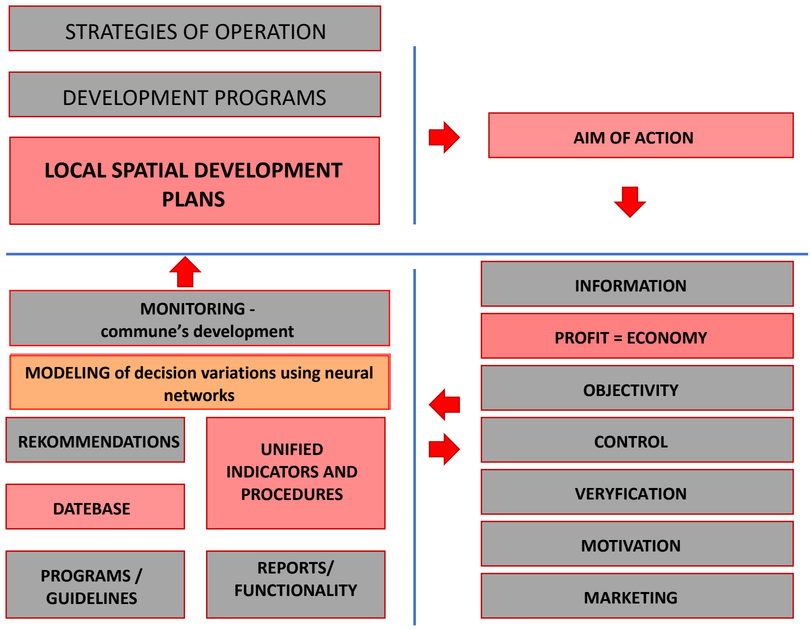

Figure 11. System to support operating a municipality

analyses of the plan arrangements variants, using a particular set of characteristics with the indicated values, gives the opportunity to present, in the case of the analysis referred to in this work, their impact on the price of the land covered by the arrangements of the plan's draft. The research results presented in this work, show that the models which use deep neural network can successfully be used in the process of e.g. analysis of variants of the same local plan draft (taking into account different parameters of the same urban indicators).

Analysis of variants of the plan allows assessing the impact of the plan arrangements already at its development stage. This is the stage which precedes a final decision of a mayor, a commune administrator, or a municipal council made about planning document records, which directly impact development of municipalities, and thus their development policy.

It must be noted that the analysis referred to in the article can become quite an effective tool in shaping development policy (Fig. 11). Unification of indicators and metrics, periodically supplied with databases information and constant monitoring of the management system, with the clearly specified objective of local government units, indicate that the modeling of decision-making variants with neural networks, can be an effective method of achieving favorable results. In the presented system for the municipality management, decisions made at the level of the local plan, relating to the municipality area and its management, have a direct impact on the space functions, its attractiveness, also in terms of investments, which translates into economic and social value.

The main limitation of the research was the input data set, in particular the number of data resulting from official registers data sets. A small number of the available historical transactions, missing values and different attributes' types (numerical, nominal and binary) caused difficulties with the preparation of appropriate input to the networks which require numeric attributes without missing values. However, a developed simple data preprocessing strategy proved to be promising. It is also known that deep neural networks provide the best predictive models if they are trained with a large number of examples (many thousands). The more promising seem the results obtained for such a limited database.

It must be assumed, regarding the research results, that using such constructed data sets, the model can be an extremely helpful tool in decision-making system e.g. in choosing the most advantageous variant, in this case, at the level of a making plan draft and establishing individual parameters and indicators identified on the basis of this document. The individual variants are possible to present in an accessible manner. Establishing the hierarchy of factors having an impact on reaching the goal, allows for translating the values of individual factors' weights to schedule changes using scenarios possible to achieve with e.g. FCM (Fuzzy Cognitive Maps).Independently, it must be accepted that the method presented in this work gives rather good results, regarding the fact that the model uses a large amount of non-numeric factors, in this case related to the characteristics of the development and the use of real estate.

References

Act (2017). Act of 27 March 2003 on spatial planning and development. Official Journal 2017, item 1073, 1566 as amended.

Act (2018). Act of 21 August 1997 on real estate management. Official Journal 2018, item 121.

Amakdouf, H., El Mallahi, M., Zouhri, A., Tahiri, A., and Qjidaa, H. (2018). Classification and recognition of 3D image of Charlier moments using a multilayer perceptron architecture. *Procedia Computer Science*, 127:226–235, doi:10.1016/j.procs.2018.01.118.

Bataineh, M. and Marler, T. (2017). Neural network for regression problems with reduced training sets. *Neural Networks*, 95:1–9, doi:10.1016/j.neunet.2017.07.018.

Bazan-Krzywoszańska, A., Mrówczyńska, M., and Tront, S. (2019). GIS technology, 3D models and mathematical models as a tool for assessing development capabilities of flood risk land to make arrangements of municipal planning documents. *Journal of Ecological Engineering*, 20(1):25–33, doi:10.12911/22998993/93866.

Cymerman, R. (2008). *Planowanie przestrzenne dla rzeczoznawców majątkowych, zarządców oraz pośredników w obrocie nieruchomościami*. Educaterra.

Cymerman, R., Grabowski, R., and Gwiaździńska, M. (1999). Wartość nieruchomości jako czynnik kreujący zagospodarowanie przestrzenne. In *Materiały VII Konferencji Naukowej Towarzystwa Naukowego Nieruchomości nt.„Wartość nieruchomości w gospodarowaniu przestrzenią"*, page 22. Wydawnictwo ART, Olsztyn.

Ertuğrul, Ö. F. (2018). A novel type of activation function in artificial neural networks: Trained activation function. *Neural Networks*, 99:148–157, doi:10.1016/j.neunet.2018.01.007.

García, N., Gámez, M., and Alfaro, E. (2008). ANN+ GIS: An automated system for property valuation. *Neurocomputing*, 71(4-6):733–742, doi:10.1016/j.neucom.2007.07.031.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press Cambridge.

Hofmann, M. and Klinkenberg, R. (2013). *RapidMiner: Data mining use cases and business analytics applications*. CRC Press Taylor and Francis Group.

Jasiński, T. and Bochenek, A. (2016). Prognozowanie cen nieruchomości lokalowych za pomocą sztucznych sieci neuronowych. *Studia i Prace WNEiZ US*, 45(1):317–327, doi:10.18276/sip.2016.45/1-25.

Juszczyk, M., Leśniak, A., and Zima, K. (2018). ANN based approach for estimation of construction costs of sports fields. *Complexity*, 2018, doi:10.1155/2018/7952434.

Łaguna, T., Łaguna, D., and Lendzin, M. (2004). Identyfikacja skutków uchwalenia i realizacji miejscowego planu zagospodarowania przestrzennego. In Łaguna, T., editor, *Ekonomiczne aspekty gospodarki przestrzennej*. Wyd. Ekonomia i Środowisko, Białystok.

Laskowska, E. (2008). Zmiana przeznaczenia gruntu a jego wartość. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie. Problemy Rolnictwa Światowego*, 5(20):42–49.

Leśniak, A. and Juszczyk, M. (2018). Prediction of site overhead costs with the use of artificial neural network based model. *Archives of Civil and Mechanical Engineering*, 18(3):973–982, doi:10.1016/j.acme.2018.01.014.

Liu, X.-s., Zhe, D., and Wang, T.-l. (2011). Real estate appraisal system based on GIS and BP neural network. *Transactions of Nonferrous Metals Society of China*, 21:s626–s630, doi:10.1016/S1003-6326(12)61652-5.

Minli, Z. and Yueran, G. (2010). Application of BP neural networks to the real estate valuation. In *Proceedings of 2010 International Conference on Construction and Real Estate Management, December 1st-3rd, 2010, Brisbane, Australia*.

Mrówczyńska, M. (2015). *Studium nad doborem metod inteligencji numerycznej do rozwiązywania problemów z geodezji inżynieryjnej*. Oficyna Wydawnicza Uniwersytetu Zielonogórskiego.

Mrówczyńska, M., Łączak, A., Bazan-Krzywoszańska, A., and Skiba, M. (2018). Improving energy efficiency with the risk of investment of reference to urban development of Zielona Góra. *Tehnički Vjesnik*, 25(3):916–922, doi:10.17559/TV-20161212120336.

Quinlan, J. R. (1992). Learning with continuous classes. In *5th Australian Joint Conference on Artificial Intelligence*, volume 92, pages 343–348. World Scientific.

Regulation (2004). Regulation of the Council of Ministers of 21 September 2004 on the property valuation and appraisal reports. Official Journal 2004, no. 207, item 2109.

Szewrański, S., Kazak, J., Żmuda, R., and Wawer, R. (2017). Indicator-based assessment for soil resource management in the Wrocław Larger Urban Zone of Poland. *Polish Journal of Environmental Studies*, 26(5):2239–2248, doi:10.15244/pjoes/70178.

Taffese, W. Z. (2007). Case-based reasoning and neural networks for real estate valuation. In *Artificial Intelligence and Applications*, pages 98–104. ACTA Press Anaheim.

Wang, X., Yan, Y., Tang, P., Bai, X., and Liu, W. (2018). Revisiting multiple instance neural networks. *Pattern Recognition*, 74:15–24, doi:10.1016/j.patcog.2017.08.026.

Xu, Q., Deng, K., Jiang, C., Sun, F., and Huang, X. (2017). Composite quantile regression neural network with applications. *Expert Systems with Applications*, 76:129–139, doi:10.1016/j.eswa.2017.01.054.

Yeh, I.-C. and Hsu, T.-K. (2018). Building real estate valuation models with comparative approach through case-based reasoning. *Applied Soft Computing*, 65:260–271, doi:10.1016/j.asoc.2018.01.029.

Zhang, Z., Ma, X., and Yang, Y. (2003). Bounds on the number of hidden neurons in three-layer binary neural networks. *Neural networks*, 16(7):995–1002, doi:10.1016/S0893-6080(03)00006-6.