The use of data by smart systems for price forecasting in the context of building customer relationships on the Lublin real estate market

A Bojanowska\(^1\) and J Lipski\(^2\)

\(^1\) Lublin University of Technology, Faculty of Management, Department of Marketing, Nadbystrzycka St. 38, 20-618 Lublin, Poland
\(^2\) Lublin University of Technology, Faculty of Management, Department of Organisation of Enterprise, Nadbystrzycka St. 38, 20-618 Lublin, Poland

a.bojanowska@pollub.pl

Abstract. This article aims to show how to design a functional application as an intelligent tool for forecasting real estate prices, among others, based on historical data. Real estate dealers or clients in this market are rarely aware of the possibility of using such a tool in the real estate valuation. This paper presents how the artificial neural network can be used for this purpose. The selected area of the Lublin real estate market has been covered by research carried out for this study. As input data for the smart system, historical data of transactions and variables characterising real estate were used, such as location of the property, its condition or size. Using the developed model, in the neural network, the sensitivity of the real estate price to selected features was carried out. It was also indicated how the use of this tool can facilitate purchasing decisions and customer relationship management on such a market.

1. Introduction
The real estate market in Poland and the world is a market with great value and high development dynamics. For example, data from the Central Statistical Office show that in the period January-November 2018, 163,800 apartments were commissioned, which is 2.5% more than last year [1]. The demand for housing and supply remain at a very high level. Customer relationship management on this market is extremely important as it is competitive and impacts not only real estate agencies but also on banks, notary offices, developers and many other important industries.

An important aspect of choosing a property by the customer is undoubtedly its price [2]. Various methods for estimating property prices can be found in the literature. Wang [3] proposed real estate price forecasting by PSO (particle swarm optimisation) and SVM (support vector machine), where PSO is chosen to determine the parameters of SVM. The real estate price forecasting cases are used to testify the forecasting performance of the proposed PSO–SVM model. Aiken [4] already points out in 2006 that trends in the real estate market can be predicted using neural networks. Other authors highlight the possibility of employing the method for spatial-temporal forecasting with an application to real estate prices [5]. Some authors have also tried to apply the theory of buyer behaviour to discover their motivation in real estate activities [6, 7]. This is a direct reference to the philosophy of customer relationship management, which is also mentioned in this article.

In this article, the goal was to design a functional application as an intelligent tool for forecasting the price of real estate, including based on historical data. The research carried out by the authors is
both literature and empirical. As part of the preparation of data for teaching the model for the purposes of this article, a survey of real estate prices was carried out together with features important to the customer, such as the area of real estate, plots, construction year, etc. For the cognitive purposes, 50 houses from the secondary and primary market located in Lublin, put up for sale on the otodom.pl portal, were randomly selected. The study was conducted in April 2019.

The computational or artificial intelligence [8], which was used in this article to achieve the goal, has already found many applications in numerous areas of engineering, data analysis, forecasting, biomedicine and others. The essence of systems based on computational intelligence is the survival and interpretation of data of a very different nature. It can be numeric, symbolic, binary, logical data or, for example, undecoded images read straight on the camera screen [9].

2. Building relationships with customers on the real estate market
Shaping customer relations and managing them has now become one of the crucial issues for entrepreneurs in almost all markets. It is not important only when the market is monopolistic or closed by special concessions and legal norms. Customers are more and more demanding [10] and are looking for newer solutions to meet their needs.

In order to effectively manage customer relations in the real estate market, CRM (Customer Relationship Management) solutions dedicated to this market can be used. Such IT solutions are built and "enclosed" with key external websites for the sale of real estate – such as, for example, Gratka, oferty.net or Vimeo [11]. The offered solutions enable intelligent search, advanced tagging of objects and dynamic lists of offer parameters. This combination creates a system that perfectly combines the two most important elements of the real estate sale process – effective service and building relationships with customers and managing the offer process. It should be remembered that CRM – or customer relationship management – is not just software. CRM is a comprehensive strategy, which, thanks to appropriate tools, strives to best meet customer needs. CRM includes, among others marketing, sales and after-sales service. These activities lead to the acquisition, maintenance and – which in the era of easy exchange of recommendations on the Internet becomes crucial – maintaining good relations with the customer also after the transaction. CRM IT systems implemented on the real estate market allow that the search for offers can be managed both by the company's representatives during business talks and – through the service – by the customers themselves. As a result, additional information about the needs and preferences of customers is gained. Users of CRM systems on the real estate market pay special attention to such features of these systems as:

- Lead/contact management system – a strong contact management platform.
- List management – as a key source of revenue, this feature is of paramount importance.
- E-mail platform – social media are used for marketing whereas the email is regarded a more professional way of communicating.
- Drip e-mail – this enables sending automatic replies such as when a user adds a newsletter subscription.
- Compatibility with other platforms – compatibility with such products as MS Office and Google Docs is a great help.
- Mobile capability – agents are frequently on the road so cloud-based CRMs compatible with all mobile devices, including both Apple and Android are in demand. It extends the control and protects sensitive data such as financials [12].

An extremely important factor when searching for and choosing a property is its price. It is this aspect of shaping the relationship that is the primary focus of this paper.

Price is one of the many factors determining consumer behaviour in various markets. It belongs to the group of marketing factors [13]. Other factors that undoubtedly influence consumer decisions are: economic (e.g. savings owned), demographic (e.g., consumer sex), psychological (e.g. personality) and socio-cultural (e.g. opinion leaders) factors. This article focuses on price because it is very important for consumers on the real estate market (it is taken into account by 67% of buyers in this market [14]. The research carried out on different markets indicated that a significant and positive
relationship exists between customers' price perceptions and their purchase intentions and that the formation of price perceptions is significantly influenced by satisfaction with pricing and services.

In the case of real estate price modelling, it typically employs models equipped with a single neuron at the ANN (Artificial Neural Network) exit. The forecasting model built in this way generates a single answer on its output. Such a solution enforces the use of MLP (Multilayer Perception) containing, among others, one hidden layer, because a model with only one learning neuron at the output would not generate the expected results. The solution proposed in this article is an alternative based on this solution.

3. Data usage by intelligent systems

In 1955, John McCarthy formulated the definition of artificial intelligence. Recently, the notion of explainable artificial intelligence has seen a resurgence, after having slowed since the burst of work on explanation in expert systems over three decades ago; for example, see Chandrasekaran et al., Swartout et al. and Buchanan and Shortliffe. He included these activities carried out by computers, whose implementation by a man would require the involvement of his intelligence. This definition is valid to this day and in light of it, intelligent systems should be understood as those offered by modern computer science, which can completely or partially replace the processes implemented by the human brain. This group of methods and IT tools includes, among others: data mining analysis systems allowing automatic detection of regularities in large data sets.

Artificial neural networks find application in recognizing and classifying patterns (assigning to category patterns), predicting time series, analysing statistical data, demining and compressing image and sound as well as in control and automation issues.

In order to execute the task set in the article, the price of real estate in the function of its essential features is predicted, and then serves to develop an application for forecasting property prices, finally a model using an artificial neural network was built. As input to the model the following features were selected:

- usable floor area of the property in square meters;
- plot area on which the property was built in square meters;
- number of rooms;
- year of construction;
- number of floors;
- primary (1) or secondary (2) market;
- nature of the building: detached (1), semi-detached (2), terraced (3);
- condition: living (1), renovation (2), finishing (3);
- the district in which the property is located:
  - City Centre (1), Szerokie (2), Ponikwoda (3), Czechow dolny (4), Nowy Slawin (5), Jakubowice Konskie (6), Kalinowka (7), Konstantynow (8), Zemborzyce (9), Botanic (10), Abramowice (11), Rudnik (12), Slawinek (13), Weglin (14), Konopnica (15), Prawiedniki (16), Slawin (17), Wrotkow (18), Marysin (19).

The above parameters of the real estate offers were taken into account as they are considered significant by the creators of many sales sites and real estate offices. These factors are also taken into account in the valuation of real estate by banks and independent experts, for instance for the purposes of preparing an appraisal report.

The research covered 50 randomly selected houses, put up for sale on a popular portal with sales and rental ads, limited to the area of Lublin. The collected data were used to build the model. The validation made use of 15% input data, whereas 15% was used for the purpose of testing.

The developed model is a black box (Figure 1) because a set of input parameters and their variability range are known, along with the corresponding set of output values representing the
transaction price for a given subset of input values. The nature of relationships between individual input features and the output of the model is unknown, as well as the nature of the impact of the interaction of individual features. Some of the features are of quantitative nature, e.g. living space, the number of rooms, etc., while the qualitative part included, for example: for living (1), for renovation (2), for finishing (3), etc. For qualitative features, numerical values have been assigned to unify the format of input data.

| Feature                                                                 | Value        |
|------------------------------------------------------------------------|--------------|
| Surface of building [m^2]                                               |              |
| Surface of building plot [m^2]                                         |              |
| How many rooms in building                                             |              |
| Year of building                                                       |              |
| How many floor                                                        |              |
| Primary market (1) or aftermarket (2)                                  |              |
| Nature of building: detached house (1), twin (2), building line (3)   |              |
| State: building to live (1), building for renovation (2), building to finish (3) |              |
| District of building                                                   |              |

**Figure 1.** The developed model as a black box.

The process of building a model using a neural network is carried out in three phases:

- defining the structure of the neural network;
- network learning;
- validation and testing.

The modelling process was carried out in the MatLab computing environment using the Toolbox Neural Networks library. The neural network learning process in the MatLab environment starts with calling the function that builds the network structure:

\[
\text{net} = \text{feedforwardnet}(\text{hiddenSizes}, \text{trainFcn})
\]  

(1)

where:

- `net` – the structure of the neural network;
- `hiddenSizes` – Row vector of one or more hidden layer sizes (default = 10);
- `trainFcn` – training function (default = 'trainlm');

the function:

\[
\text{net} = \text{train}([\text{net}, \text{x}, \text{t}])
\]  

(2)

launches the network learning phase on the input data contained in matrix `x`. This matrix has the number of rows equal to the number of input data and the number of columns equal to the number of the training stream. The `t` parameter is the property price values corresponding to the next input data combinations. The structure of the network constructed in this way is shown in Figure 2. Designed a feed-forward network includes an input layer, a hidden layer of 10 neurons and one output.

**Figure 2.** Neural network structure.
As a result of learning, the following model parameters shown in the regression graphs of Figure 3 were obtained.

Figure 3. Network learning results on the training data set. Regression R-Values measure the correlation between outputs and targets. R-value of 1 means a close relationship, 0 a random relationship.

On the basis of the results of the regression analysis, it can be stated that the model largely maps the relationship between the price of real estate and the characteristics selected for research. The value of the regression coefficient $R = 0.88$ in the testing process allows to accept that the model maps the relationship between price and property features with acceptable accuracy. The model obtained in this way was used in an application developed in the MatLab computing environment. This application allowed for forecasting real estate prices on the basis of defined parameters consistent with the character with the model being taught. A number of studies on the sensitivity of the price to changes in selected parameters of the facility were also conducted. The results are shown in Figures 4 and 5.
Figure 4. The nature of changes in the price of real estate from the usable area for the set other parameters.

Figure 5. The dependence of the real estate price on the construction year, assuming other features at a constant level.

Using the forecasting model in the neural network, the calculator of the predicted property price for the introduced parameters was developed. The user interface of the application is shown in Figure 6. After selecting the property features and entering the desired value or option number into the text field, the simulation process starts and displays the expected value of the property.
Figure 6. The interface of the real estate price calculator determined as a result of simulation in the learned neural network.

4. Conclusions

The conducted research on the use of the neural network for modelling causal relationships between property features and their price confirmed the effectiveness of the method. The process of collecting data for network learning was limited to a single area and selected districts. Therefore, the use of the learned network applies to the city of Lublin. The calculator developed using the model immediately estimates the price of the property with an accuracy rate of around 80 to 85%. The training data did not take into account the time of the transaction, because the interest rate on loans was stable and the inflation was around 2%. Modelling was carried out, based on data from 2019, in the MATLAB R2018b computing environment using Toolbox Neural Net Fitting – Fit data by training and two-layered-forward network. The MATLAB GUIDE (Graphic User Interface Design Environment) programming environment was used for the software of the "Real estate price calculator" application. The developed calculator allows you to analyse the impact of the value of individual property features on the expected price. This improves the experience of a potential customer, in the conditions of limiting the investment budget, to make a decision regarding the allowed changes in the area of parameters of the purchased object. Many properties of real estate, as shown by simulation analyses, are related to the price of non-linear relationships. This means that the selection of seemingly similar available offers may differ significantly in price with relatively small changes in selected parameters such as, for example, the size of the plot, the number of rooms or usable space. The calculator facilitates optimisation of the property selection. It can be useful for customers but can also support the process of customer relationship management for many entities operating within the real estate market. Offering such a tool to customers will certainly encourage them to cooperate and open up new areas for building relationships with him. It can also be used to collect data on customer preferences and encourage them to take self-service activities. The solution proposed in the article may operate independently or constitute an extension to the CRM software module.

The neural network used to build the model is an intelligent system applied in the developed application. Modelling and research undertaken in this article are set for further development, for example using more input for the model. They can also be the basis for building more effective CRM IT tools for the real estate market.
5. References

[1] Real estate sales in 2018. Statistics Poland

[2] Ariyawansa R G 2009 Housing Market: A Review of Purchase Decision of Potential Buyers, Department of Estate Management and Valuation (Sri Lanka: Faculty of Management Studies and Commerce University of Sri Jayewardenapura) doi: 10.4038/besl.v8i1.1906

[3] Wang X, Wen J, Yhag Z and Wang Z 2014 Real estate price forecasting based on SVM optimized by PSO Optik 125(3) doi: 10.1016/j.ijleo.2013.09.017

[4] Aiken M and Bsart M 2006 Forecasting Market Trends with Neural Networks Information Systems Management 16(4) doi: 10.1201/1078/43189.16.4.19990901/31202.6

[5] Pace R K, Barry R, Gilley O W and Sirmans C F 2000 A method for spatial-temporal forecasting with an application to real estate prices International Journal of Forecasting 16(2) doi: 10.1016/S0169-2070(99)00047-3

[6] Chia J, Harun A, Kassim A W M, Martin D and Kepal N 2016 Understanding factors that influence house purchase intention among consumers in Kota Kinabalu: An application of buyer behavior model theory Journal of Technology Management and Business 3(2)

[7] Zeng R 2013 Attributes influencing home buyers' purchase decisions: a quantitative study of the Wuhan residential housing market (DBA thesis, Lismore NSW: Southern Cross University)

[8] Miller T 2019 Explanation in artificial intelligence: Insights from the social sciences Artificial Intelligence 267 doi: 10.1016/j.artint.2018.07.007

[9] Rutkowski L 2019 Metody sztucznej inteligencji (Warszawa: Wydawnictwo Naukowe PWN)

[10] Hyken S 2012 Today's Customers Demand Customer Service On Their Terms https://www.forbes.com/sites/shephyken/2016/11/12/todays-customers-demand-customer-service-on-their-terms/ (12.09.2019)

[11] Real Estate CRM Platform https://makolab.com/pl-pl/rozwiazania/dedykowane-rozwiazania-biznesowe/platfom-crm-dedykowana-bran%C5%BCy-nieruchomo%C5%9Bi (13.05.2019 in:Polish)

[12] Taren J 2019 Top features for real estate CRM https://www.discovercrm.com/real-estate-crm-features.html (20.08.2019)

[13] Wawrzyniak A 2015 Modelowanie wieloagentowe jako narzędzie do badania zachowania konsumentów Zeszyty Naukowe SGGW w Warszawie Polityki Europejskie Finanse i Marketing 14(63) 134-46 doi:10.7172.1644-9584.38.7

[14] Cena, koszty utrzymania i bezpieczeństwo najważniejsze przy zakupie mieszkania https://www.wirtualnemedia.pl/artykul/najwazniejsze-warunki-przy-zakupie-mieszkania (23.05.2019 in:Polish)

[15] Munnukka J 2008 Customers' purchase intentions as a reflection of price perception Journal of Product & Brand Management 17(3) 188-96 doi: 10.1108/10610420810875106

[16] Jasioński T and Bochenek A 2016 Prognozowanie cen nieruchomości lokalowych za pomocą sztucznych sieci neuronowych Studia i Prace WNEiZ US 45(1) 317-27 doi: 10.18276/sip.2016.45/1-25

[17] McCarthy J 2007 What is Artificial Intelligence? Computer Science Department http://www-formal.stanford.edu/jmc/whatisai/ (12.09.2019)

[18] Chandrasekar B, Tanner M C and Josephson J R 1989 Explaining control strategies in problem solving IEEE Expert 4(1) 9-15 doi: 10.1109/64.21896

[19] Swartout W R and Moore J D 1993 Explanation in second generation expert systems Second Generation Expert Systems eds J M David, J P Krivine and R Simmons (Berlin, Heidelberg: Springer) pp 543-85 doi: 10.1007/978-3-642-77927-5_24

[20] Buchanan B and Shortliffe E 1984 Rule-based Expert Systems: the MYCIN Experiments the Stanford Heuristic Programming Project (Addison-Wesley) doi:10.1016/0004-3702(85)90067-0

[21] Holyński M 1979 Sztuczna inteligencja (Warszawa: Wiedza Powszechna)

[22] Rojas R 2013 Neural Networks: A Systematic Introduction (Berlin, Heidelberg: Springer)

[23] Tadeusiewicz R 1993 Sieci neuronowe (Warszawa: Akademicka Oficyna Wydawnicza)