

## ANALYSIS OF DIFFERENT APPROACHES TO REAL ESTATE APPRAISAL

Vladimir Surgelas, Irina Arhipova, Vivita Pukite

Latvia University of Life Sciences and Technologies, Latvia

dr.engenho@gmail.com

### Abstract

The traditional valuation of real estate in the field of civil engineering did not include the uncertainty of human behaviour, which cannot be explained by the traditional approach. There are different valuation methods for real estate appraisal, which are basically classified into three groups as a classic, statistical and advanced. In this article, we estimated the different housing price models using the sample of 37 residential apartments in Riga, Latvia, October 2018. In order to evaluate if there is a possible association between the variables involved in relation to the property price, the analytical data were analysed by correlation analysis, analysis of variance (ANOVA), regression analysis, covariance analysis (ANCOVA), principal component analysis (PCA) and cluster analysis. The models estimation results show that using ANCOVA models for the prices forecasting the model fitting to data is less than 58%. The preliminary results of this study suggest that the estimated properties can be distributed in 4 groups, depending on number of rooms, area and age. In addition, the decision tree was created based on algorithms (J48) and a preliminary definition of the best rules was made. The decision tree presents an accuracy of 84% with 31 accepted instances for a total of 37 currently classified instances.

**Key words:** regression analysis, civil engineering, fuzzy logic, decision tree, appraisal real estate.

### Introduction

Real estate appraisals are very important and are especially necessary for asset valuation, property taxes, insurance estimates, sales or rent transactions, risk analysis, estate planning, litigation and are relevant to government or investor decision making. Housing prices depend not only on various economic indices but also on the impact of other factors such as natural disasters, wars, market speculation, environmental variables (Aderibigbe & Chi, 2018). There are different valuation methods for real estate appraisal (Yeh & Hsu, 2018), however, in some types of real estate appraisals, a problem arises in determining real estate values.

For example, using the method of linear or nonlinear multiple regression, the problem of multicollinearity is actual, when predictor variables in the model are correlated with other predictor variables (Wheeler & Tiefelsdorf, 2005). (Dantas *et al.*, 2007) clarifies that there may be a problem of price estimation in real estate in the case of the use of traditional econometric methods, because the results tend to be biased, inefficient or inconsistent, given the neglect of spatial effects in the data. The same authors proposed a study considering the use of spatial econometrics. In this study, it was concluded that positive spatial autocorrelation in the data, although the negotiations of real estate purchase does not occur independently, as considered in the model of hedonic prices, there is an interaction between these prices, so that a negotiation of a real estate/apartment for a high price will generate a growing influence on the prices of neighbouring properties.

On the other hand (Wheeler *et al.*, 2014), it makes clear that in the case of spatial structure choice, this affects the interpretation of the parameters for the variables with which it is correlated, that is, it is also

a type of multicollinearity. Therefore, the noncritical use of spatial econometrics may cause problems in the interpretation of individual parameters. Another possible solution would be to weaken multicollinearity by constructing Bayesian models that use variable coefficient processes to model non-constant linear relationships between variables (Gargallo, Miguel, & Salvador, 2017). According to (Yeh & Hsu, 2018) the traditional statistical method, basically, the type and measure of the variable define the most suitable model for use. The traditional valuation model is based on the hedonic model to estimate housing prices (Gargallo, Miguel, & Salvador, 2017).

On the other hand, the advanced method (Smętek & Trawiński, 2011) is related to fuzzy logic (Zadeh, 1978), artificial intelligence, decision tree allied to artificial learning algorithms (Eibe, Hall, & Witten, 2016). Although there is a limited number of studies on fuzzy logic for the real estate industry, there is a strong tendency for improvements related to the process of knowledge representation through fuzzy logic, supported by computational tools (Hüllermeier, 2015). In this case, when there are scattered data, the fuzzy model demonstrates a capacity for analytical compatibility with the problem (Zadeh, 1978). Moreover, fuzzy logic contemplates the subjectivities of the variables involved, the imprecision and uncertainty of human expression. Fuzzy logic is due to the need to consider subjective, ambiguous, and imprecise aspects that are difficult to measure to determine the value of a property. Nevertheless, the use of fuzzy logic is an alternative, there are observed factors that may be slowing the progress of this implementation in the area of real estate valuation. This fact is related to the number of fuzzy rules generated by the model. These rules grow geometrically, directly proportional to the number of attributes selected for

property valuation. However, this does not seem to be a problem, but a solution.

In this way and without losing the scientific character of the traditional evaluation, the proposal follows with the help of the decision tree allied to the algorithms of artificial learning. Accordingly (Eibe, Hall, & Witten, 2016), the association generates rules that describe the most relevant patterns present in the data. These rules are composed of precedents and consequents, this choice of rules promotes a discovery of patterns of behaviour. This behaviour will indicate that the proposed fuzzy model will present behaviour in the data processing of the two groups, and standard deviation and deviation percentage within a similar range. Therefore, there are some situations in practice where the option to adopt the hedonic method, so to speak, binary logic is not always the best solution to determine the market value of a property. Because there is a possibility of appearance of disagreement in the fullness of estimating some real estate prices. In addition, each country has different cultural and socio-economic structures. This particularity results in the difficulty of defining a world standard. But from this, possible price misunderstandings should be avoided.

Thus, an important point was the definition of real estate value determined by the International Council of Evaluation Standards (IVSC), whose definition reaches international acceptance. In addition, the need to seek a more palpable solution to the process of evaluating buildings is exacerbated by daring projects. In addition, changes in the construction production chain, including products, suppliers, materials and a process can significantly improve environmental efficiency and the economy during construction used in life cycle assessment (Surgelas, Marques, & Rodrigues, 2010). Thus, according to Baumane (2011) to implement an evaluation process, the object of this evaluation must be clearly defined, then choose the evaluation method. This research advances in the study of the problem and seeks an alternative for the forecasting of prices in the evaluation of real estate. Although there is a traditional method and literature deals with such techniques, there is controversy over which one would bring the best performance and with a greater degree of facility and clarity. Therefore, this controversy is due to the difficulty in finding adequate predictors because they often do not fit into the binary model.

The aim of the article is estimating different housing price models using the sample of 37 residential apartments in Riga, Latvia, October 2018 in order to evaluate if there is a possible association between the variables involved in relation to the property price. Furthermore, this article is part of the thesis in progress and is structured in 3 step: – Step 1, sample collection and statistical treatment based

on binary logic. – Step 2, decision tree, and step 3, a creation of the fuzzy model with analysis validation of the results.

In the first step, the present study opted for the hedonic method to estimate apartments' prices, since this model is the technique most used by civil engineering professionals. Due to the emergence of multicollinearity, we have chosen to find more adequate regression models that relate the response variables to the variables involved in the property valuation analysis. Thus, property prices based on the physical characteristics and location of the property were analysed through correlation analysis, ANOVA analysis, regression analysis, covariance analysis (ANCOVA), analysis of major components (PCA) and cluster analysis. To evaluate the market price of the appraised apartment. In order to evaluate the price forecasting accuracy, the following three distance metrics are used: root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percent error (MAPE). While both MAE and RMSE are simply measures of discrepancies between the predicted values and the actual observations, MAPE measures scaled discrepancies at each time interval (Ahn *et al.*, 2012). Finally, a decision tree based on algorithms and preliminary definition of the best rules (Tobergte & Curtis, 2013) was created. The hedonic model was used in this research to estimate housing prices, since this method is considered the most used (Gargallo, Miguel, & Salvador, 2017) and incorporated in the international standards of evaluation (IVS) and in the European evaluation standards (EVS) as an approach that estimates a market value. This method has been criticized due to the problems of multicollinearity in the estimation of the parameters, which are due to the very similar characteristics of the houses in the same area, which makes it difficult to estimate the regression coefficients (Bárcena *et al.*, 2014).

## Materials and Methods

This article is related to improvements in the real estate valuation process. In this first step the different housing price models are estimated using the sample of 37 residential apartments in Riga, Latvia, October 2018 in order to evaluate if there is a possible association between the variables involved in relation to the price of the property. The selection of random samples occurred through a direct comparison of market data of similar elements in relation to the intrinsic and extrinsic characteristics of the property, and availability in the worldwide computer network. As for the research structure, the variables that are, in principle, relevant for explaining the formation of values and the supposed relations between them and the dependent variable are chosen. In relation to the quality of the data surveyed, this study opted

for: Latvian Government Office, <http://liaa.gov.lv/en/business-latvia/real-estate-market-research>; Colliers International Group Inc. State Land Service; The Real Estate Company, <http://www.arcoreal.lv/en>; others, such as <https://city24.lv>, and experience of own researcher in the formation of value. In this first step stage, the present study opted for the hedonistic method to estimate housing prices, since this model is the technique most used by civil engineering professionals. Based on existing theories, acquired knowledge, common sense and other attributes, the dependent variable  $Y$  is housing price (EUR m<sup>2</sup>) and following factors or independent variables were chosen: an area (m<sup>2</sup>), apparent age (new(1), renew(2), old(3)), apartment conservation (new(1), renew(2), old(3)), building conservation (new(1), renew(2), old(3)), number of rooms (2) or (3), car parking (inside build(1), outside build(2), no parking(3)), sunlight = insolation (direct(1), not direct(2)), heating inside the building (individual(1), share(2)), environmental variable or park\_river (near(1), far(2)).

The research design consists of the following steps:

1. check the qualitative factors significance for housing price by one-way ANOVA model,
2. check the quantitative factor of area (m<sup>2</sup>) significance for housing price by simple regression analysis,
3. check the ordinal scale factors multicollinearity by Spearman correlation analysis,
4. check both the quantitative factor of area (m<sup>2</sup>) and qualitative factors significance for housing price by ANCOVA model,
5. get the new uncorrelated principal components from correlated factors by PCA,
6. using the new uncorrelated principal components check the housing price model significance by regression analysis,
7. evaluate the price forecasting accuracy using mean absolute percent error (MAPE),
8. define similar groups of real estates by cluster analysis,
9. check the housing price model using the decision tree model based on algorithms (J48-data mining algorithms) and define the best association rules.

The present study uses the variable price ( $Y$ ) and area logarithmic transformation to avoid possible misunderstandings in the results. The decision to transform dependent variable into  $LnY$  is made to test the dependent variable  $Y$  in the absolute term (Euro m<sup>2</sup>), to change in the relative term (%). This is because the use of logarithms of dependent or independent variables may allow nonlinear relations between the explained variable and the explanatory variables. Thus, estimates using logs are less sensitive to unequal (or extreme) observations because of

the considerable narrowing that may occur in the amplitude of variable values. Thus, the test result may indicate whether to reject or not to reject the null hypothesis that the data comes from a normally distributed population.

### Results and Discussion

The limitations of this model are presented, by different models of housing prices were found using samples from 37 residential apartments in Riga, Latvia. To determine if any of the differences between the means for different price groups, at the 1<sup>st</sup> stage the one-way ANOVA was made, show, that the factors such as apparent age, apartment conservation, building conservation, car parking, sunlight (insolation), heating inside the building and environmental variable are significant with  $P=95\%$ , but the factor like the number of rooms is significant with  $P=85\%$ , for the models, where dependent variable is  $Y$  or  $LnY$ . At the 2<sup>nd</sup> stage the simple regression analysis was made to check if the area factor is significant or not for the housing price using logarithmic transformation. As a result, it was concluded that the area factor is significant with  $P=95\%$ , but the determination coefficient  $R^2=10.2\%$  shows, that the model cannot be used for price forecasting. From previous data analysis it can be concluded, that all factors are significant, but the one-factor models are not enough for price estimation. So, the next step is the model transformation or additional factors should be putted to the model. At the 3<sup>rd</sup> stage the factors such as apparent age, apartment conservation, building conservation, number of rooms, car parking, sunlight or (insolation), heating inside the building and environmental variable were checked to the problem of multicollinearity using Spearman correlation analysis (Table 1). All factors correlate among themselves, except rooms, so there is the problem of multicollinearity and it is not possible to put these factors together, except rooms, at the same model.

At the 4<sup>th</sup> stage the quantitative factor of area (m<sup>2</sup>) and two qualitative factors and its interaction effect significance for housing price by ANCOVA model were checked, where one of the qualitative factors is number of rooms (Table 2). The best model includes the factors of area (m<sup>2</sup>), number of rooms, apparent age, interaction effects of number of rooms  $\times$  apparent age and area (m<sup>2</sup>)  $\times$  apparent age, but the determination coefficient  $R^2=58\%$  show, that model cannot be used for price forecasting. The decision is that: the best model is  $Y: \text{Area} + \text{Room} + \text{Age} + \text{Room} \times \text{Age} + \text{Area} \times \text{Age}$  or  $LnY: Ln\text{Area} + \text{Room} + \text{Age} + \text{Room} \times \text{Age} + Ln\text{Area} \times \text{Age}$ .

Figure 1 illustrates the ANCOVA graphical analysis, for example, the price: rooms + age shows

Table 1

**Spearman rho, correlation analysis**

variables	age	consv ap	consv build	rooms	parking_1	insolation	park_river
consv ap	0.822 0.000	–	–	–	–	–	–
consv build	0.507 0.001	0.697 0.000	–	–	–	–	–
rooms	0.180 0.285	-0.067 0.692	-0.239 0.154	–	–	–	–
parking_1	0.870 0.000	0.819 0.000	0.456 0.005	0.153 0.367	–	–	–
insolation	0.064 0.707	0.143 0.397	0.291 0.081	-0.175 0.299	0.072 0.671	–	–
park_river	0.198 0.240	0.358 0.029	0.291 0.081	-0.175 0.299	0.175 0.301	0.630 0.000	–
heating	0.470 0.003	0.469 0.003	0.436 0.007	-0.017 0.920	0.388 0.018	0.549 0.000	0.549 0.000

Cell Contents: Spearman rho  
P-Value

Table 2

**The ANCOVA analysis**

Model	Area	Room	Age	Room × Age	Area × Room	Area × Age	Determination coefficient (%)
Y: Area + Room + Age + Room × Age + Area × Room + Area × Age	0.584	0.346	0.010	0.000	0.614	0.017	58.61
Y: Area + Room + Age + Room × Age + Area × Age	0.748	0.065	0.009	0.000	–	0.015	58.21
LnY: LnArea + Room + Age + Room × Age + LnArea × Room + LnArea × Age	0.573	0.747	0.078	0.005	0.815	0.084	53.11
LnY: LnArea + Room + Age + Room × Age + LnArea × Age	0.599	0.155	0.051	0.004	–	0.057	53.01
Y: Area + Room + Parking + Room × Parking + Area × Room + Area × Parking	0.311	0.715	0.330	0.431	0.491	0.470	30.10
Y: Area + Room + Parking	0.274	0.709	0.047	–	–	–	24.52
LnY: LnArea + Room + Parking + Room × Parking + LnArea × Room + LnArea × Parking	0.432	0.917	0.334	0.395	0.851	0.370	33.73
LnY: LnArea + Room + Parking	0.266	0.637	0.033	–	–	–	27.43

that for an apartment with 2 and 3 rooms and with a new age (1) the price of the apartment will decrease as the number of rooms increases. However, for an apartment in the renovated situation (2) the process is reversed. Thus, for an apartment with 3 renovated rooms the price of this property reaches a reasonable amount compared to the apartment with 2 renovated rooms. And in the situation where both are old (3) the price difference is not much different.

At the 5<sup>th</sup> stage the Principal Component Analysis (PCA) was made for correlated 8 factors of apparent

age, apartment conservation, building conservation, number of rooms, car parking, sunlight, heating inside the building and environmental variable to get new uncorrelated principal components PC for the next regression analysis. The total explained data variance, using 4 new uncorrelated factors is 81.4% (Table 3). The first PC can be explained as *comfortability* factor, which is highly correlated to the factors of car parking, apparent age and apartment conservation. The second PC can be explained as *insolation* factor, which is highly correlated to the factor of sunlight (or

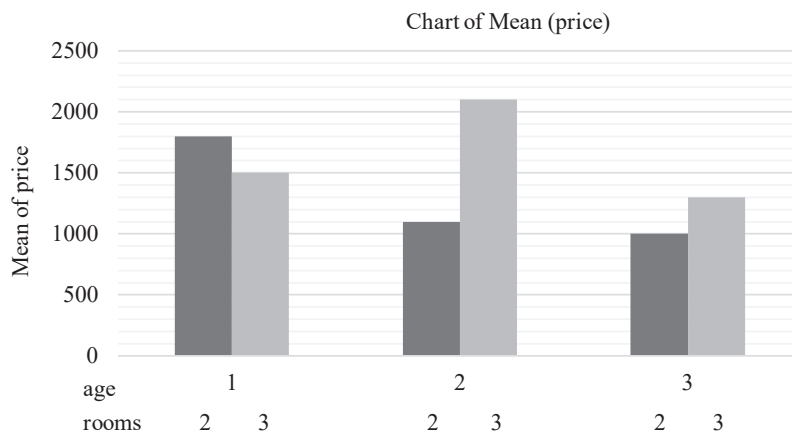


Figure 1. ANCOVA analysis graphs.

Table 3

**PCA analysis – Sorted Rotated Factor Loadings and Communalities**

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
parking_1	0.941	0.000	0.000	0.000	0.000	0.000	0.000	0.000
age	0.892	0.000	0.000	0.000	0.000	0.000	0.000	0.000
consv ap	0.803	0.000	0.000	0.000	0.000	0.000	0.000	0.000
insolation	0.000	0.916	0.000	0.000	0.000	0.000	0.000	0.000
rooms	0.000	0.000	0.986	0.000	0.000	0.000	0.000	0.000
park_river	0.000	0.000	0.000	0.907	0.000	0.000	0.000	0.000
consv build	0.410	0.000	0.000	0.000	-0.874	0.000	0.000	0.000
heating	0.000	0.000	0.000	0.000	0.000	-0.867	0.000	0.000
Variance	2.5998	1.0428	1.0372	1.0325	1.0282	0.9523	0.1571	0.1501
% Var	0.325	0.130	0.130	0.129	0.129	0.119	0.020	0.019
CumulativeVar	0.325	0.455	0.685	0.814	-	-	-	-

Table 4

**Factor Score Coefficients - Formulas to calculate new uncorrelated factors**

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
age	0.472	0.066	-0.081	-0.031	0.220	0.135	-0.782	2.084
consv ap	0.205	0.062	-0.008	-0.085	0.153	0.087	2.423	-0.332
consv build	-0.297	-0.108	0.164	0.035	-1.366	0.118	-0.675	-0.093
rooms	-0.139	0.056	1.068	0.097	-0.242	0.062	0.236	-0.161
parking_1	0.647	0.043	-0.080	-0.035	0.247	0.125	-1.019	-1.574
insolation	0.064	1.319	0.048	-0.388	0.121	0.324	0.318	0.225
park_river	-0.072	-0.372	0.079	1.316	-0.043	0.233	-0.483	0.066
heating	-0.186	-0.271	-0.036	-0.205	0.112	-1.384	0.098	-0.253

Factor 1: parking\_1, age and consv\_ap (comfortability) = 0.472×age + 0.205×consv\_ap - 0.297×consv\_build - 0.139×rooms + 0.647×parking\_1 + 0.064×insolation - 0.072×park\_river - 0.186×heating;

Factor 2: insolation = 0.066×age + 0.062×consv\_ap - 0.108×consv\_build + 0.056×rooms + 0.043×parking\_1 + 1.319×insolation - 0.372×park\_river - 0.271×heating;

Factor 3: rooms = - 0.081×age - 0.008×consv\_ap + 0.164×consv\_build + 1.068×rooms - 0.080×parking\_1 + 0.048×insolation + 0.079×park\_river - 0.036×heating;

Factor 4: park\_river = - 0.031×age - 0.085×consv\_ap + 0.035×consv\_build + 0.097×rooms - 0.035×parking\_1 - 0.388×insolation + 1.316×park\_river - 0.205×heating.



Table 5

Summary Model description and determination coefficient R<sup>2</sup>

Models	R <sup>2</sup> (%)
• Three-factors ANCOVA	-----
Y: area + room + age + rooms×age + area×age	58.21
LnY: LnArea + rooms + age + rooms×age + LnArea×age	53.01
• Two-factors ANCOVA	-----
Y: area + parking_1	24.18
LnY: LnArea + parking_1	26.92
• Two-way ANOVA	-----
Y: rooms + parking_1	21.59
LnY: rooms + parking_1	24.53
• Regression models with PCA	-----
Price (Y): F1 + F2 + F3 +F4	29.98
LnPrice (LnY): F1 + F2 + F3 +F4	34.47
Price (Y): area + F1 + F2 + F4	29.79
LnPrice (LnY): LnArea + F1 + F2 + F4	34.87

insolation), the third PC can be explained as *rooms' factor*, which is highly correlated to the factor of number of rooms and the fourth PC can be explained as *environmental factor*, which is highly correlated to the factor of park\_river. And the formulas to calculate new uncorrelated factors are shown in (Table 4).

At the 6<sup>th</sup> stage the multiple regression models were analysed with new four uncorrelated PC (comfortability, insolation, rooms and environment) and fifth factor of area. To avoid the problem of multicollinearity, the Pearson correlation analysis was made and concluded, that area factor is correlated with rooms' PC (*r*=0.755). Thus, the following 4 models were analysed:

- *LnY* or *Y* (housing price) depending on comfortability, insolation, rooms and environment,
- *LnY* or *Y* (housing price) depending on area, comfortability, insolation and environment,

but the determination coefficient for all models is less than 35% and canno't be used for price forecasting.

Moreover, the table 5 shows the following a summary for ANCOVA; two-way ANOVA; Regression analysis with PCA. The decision model description: the most appropriate model is:

Models: Price Y: area + room + age + room×age + area×age with determination coefficient R<sup>2</sup>(58%).

The Regression Equation results:

rooms	age	equation
2	1	price = 739 + 17.2 area
2	2	price = 2213 – 23.83 area
<b>2</b>	<b>3</b>	<b>price = 430 + 12.58 area</b>
3	1	price = 277 + 17.2 area
3	2	price = 4069 – 23.83 area
3	3	price = 295 + 12.58 area

At the 7<sup>th</sup> stage the price forecasting accuracy using mean absolute percent error (1) was made.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left( \frac{Y_t - \hat{Y}_t}{Y_t} \right)^2 \times 100 \quad (1)$$

For model forecasting evaluation MAPE = 10%, and according (Ahn *et al.*, 2012) it means that on average the difference between actual and theoretical model is 10% (reasonable and within the normality of this case study). Then, at this stage of the analysis, it is possible to propose an example to quantify the value of the square meter of the appraised apartment.

At the 8<sup>th</sup> stage similar groups of real estates, using factors of area, room and age were found by cluster analysis for 37 real estates: 1<sup>st</sup> cluster consists of real estates with average area in 88.9 m<sup>2</sup>, 3 rooms and age renew (2) or old (3); 2<sup>nd</sup> cluster consists of real estates with average area in 70.8 m<sup>2</sup>, 3 rooms and age new (1) or renew (2); 3<sup>rd</sup> cluster consists of real estates with average area in 54.5 m<sup>2</sup>, 2 rooms and age new (1) or renew (2); 4<sup>th</sup> cluster consists of real estates with average area in 41.2 m<sup>2</sup>, 2 rooms and age renew (2) or old (3).

At the last stage the similar groups a decision tree based on algorithms (J48) and preliminary definition of the 10 best and the association rules were preliminarily chosen by algorithmic 'apriori' with accordance 83.78% (84%):

1. Price\_category=1 16 → park\_river=2 16 <conf: (1)> lift: (1.48) lev: (0.14) [5] conv: (5.19)
2. Price\_category=1 16 → heating=2 16 <conf: (1)> lift: (1.32) lev: (0.11) [3] conv: (3.89)
3. heating=2 Price\_category=1 16 → park\_river=2 16 <conf: (1)> lift: (1.48) lev: (0.14) [5] conv: (5.19)

4.  $\text{par\_river}=2$   $\text{Price\_category}=1$   $16 \rightarrow \text{heating}=2$   
 $16$  <conf: (1)> lift: (1.32) lev: (0.11) [3] conv:  
(3.89)
5.  $\text{Price\_category}=1$   $16 \rightarrow \text{park\_river}=2$   $16$  heating  
 $=2$   $16$  <conf: (1)> lift: (1.61) lev: (0.16) [6] conv:  
(6.05)
6.  $\text{parking\_1}=3$   $\text{heating}=2$   $15 \rightarrow \text{age}=3$   $15$  <conf:  
(1)> lift: (2.18) lev: (0.22) [8] conv: (8.11)
7.  $\text{insolation}=2$   $\text{Price\_category}=1$   $15 \rightarrow \text{park\_river}=2$   
 $15$  <conf: (1)> lift: (1.48) lev: (0.13) [4] conv:  
(4.86)
8.  $\text{insolation}=2$   $\text{Price\_category}=1$   $15 \rightarrow \text{heating}=2$   $15$   
<conf: (1)> lift: (1.32) lev: (0.1) [3] conv: (3.65)
9.  $\text{insolation}=2$   $\text{heating}=2$   $\text{Price\_category}=1$   $15 \rightarrow$   
 $\text{park\_river}=2$   $15$  <conf: (1)> lift: (1.48) lev: (0.13)  
[4] conv: (4.85)
10.  $\text{insolation}=2$   $\text{park\_river}=2$   $\text{Price\_category}=1$   $15$   
 $\rightarrow \text{heating}=2$   $15$  <conf: (1)> lift: (1.32) lev: (0.1)  
[3] conv: (3.65)

### Conclusions

The limitations of this model are presented by different models of housing prices that were found using samples from 37 residential apartments in Riga, Latvia. This study found the existence of an association between the variables involved in relation to the price of the property. Then, after the entire analysis process as the result the most appropriate ANCOVA model was defined, where three factors (property area in m<sup>2</sup>, number of rooms, property age as old, renew or new) and two factors interactions effect (number of rooms  $\times$  property age and property area  $\times$  property age) are included in the model. The estimation of the model shows that the coefficient of determination R<sup>2</sup> is 58% of the price change was

explained by the property price forecast using the factors included in the model. The preliminary results of this study suggest that the estimated properties can be distributed by 4 groups: renew or old properties with average area 88.9 m<sup>2</sup> and 3 rooms; new or renew properties with average area 70.8 m<sup>2</sup> and 3 rooms; new or renew properties with average area 54.5 m<sup>2</sup> and 2 rooms; renew or old properties with average area 41.2 m<sup>2</sup> and 2 rooms. The association rules were generated by the algorithm that preliminarily listed 10 best rules. The decision tree presents an accuracy of 84% with 31 accepted instances for a total of 37 currently classified instances, and finally, the third and final step is to establish a fuzzy logic model using the 10 best association rules created by the *apriori* algorithm. algorithms for learning/data mining), according to the results of the regression equation described above. In order of hand, the purpose of this article is to estimate the different models of the price of housing and, for example, if you have an apartment with rooms = 2 and age = 3 and area = 45 m<sup>2</sup>, then the theoretical price is  $Y = 430 + 12.58 \times 45\text{m}^2$ . In step 1, it is possible to say that of this estimated value (Y) there is an acceptable margin of 10% for negotiating the apartment. And finally, the relationship involving human behavior in the valuation of the property will be applied in the last step 3 and will approach the Fuzzy Logic to estimate the market price of the apartment and thus compare with the values obtained in the traditional binary technique. The quality of the evaluation will depend on the quality of the information available to prepare it.

### Acknowledgements

Research is supported by Latvian State Scholarship.

### References

1. Aderibigbe, T., & Chi, H. (2018). Investigation of Florida Housing Prices using Predictive Time Series Model. In *Proceedings of the Practice and Experience on Advanced Research Computing – PEARC '18* (pp. 1–4). New York, New York, USA: ACM Press. DOI: 10.1145/3219104.3229253.
2. Ahn, J.J., Byun, H.W., Oh, K.J., & Kim, T.Y. (2012). Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. *Expert Systems with Applications*, 39(9), 8369–8379. DOI: 10.1016/j.eswa.2012.01.183.
3. Baumane, V. (2011). *Improvement of cadastral assessment models*. Unpublished doctoral dissertation, Latvia University of Agriculture Faculty of Economic. Jelgava, Latvia.
4. Bárcena, M.J., Menéndez, P., Palacios, M.B., & Tusell, F. (2014). Alleviating the effect of collinearity in geographically weighted regression. *Journal of Geographical Systems*. DOI: 10.1007/s10109-014-0199-6.
5. Dantas, R.A., Magalhães, A.M., & Vergolino, J.R. de O. (2007). Avaliação de imóveis: a importância dos vizinhos no caso de Recife (Valuation of Real Estate: The importance of the neighbours in the case of Recife). *Economia Aplicada*, 11(2), DOI: 10.1590/S1413-80502007000200004. (in Portuguese)
6. Eibe, F., Hall, M.A., & Witten, I.H. (2016). The WEKA Workbench. Online Appendix for ‘Data Mining: Practical Machine Learning Tools and Techniques’, Morgan Kaufmann, Fourth Edition, 2016.
7. Gargallo, P., Miguel, J.A., & Salvador, M.J. (2017). MCMC Bayesian spatial filtering for hedonic models in real estate markets. *Spatial Statistics*, 22, 47–67. DOI: 10.1016/j.spasta.2017.07.010.
8. Hüllermeier, E. (2015). From knowledge-based to data-driven fuzzy modeling: Development, criticism, and alternative directions. *Informatik-Spektrum*. DOI: 10.1007/s00287-015-0931-8.

9. Smętek, M., & Trawiński, B. (2011). Selection of heterogeneous fuzzy model ensembles using self-adaptive genetic algorithms. *New Generation Computing*. DOI: 10.1007/s00354-010-0305-3.
10. Surgelas, F.M.A., Marques, G.F., & Rodrigues, C.D.S. (2010). Life cycle analysis of ceramic versus painting materials applied to external walls. *International Journal for Housing Science and Its Applications*.
11. Tobergte, D.R., & Curtis, S. (2013). *Data Mining – Practical Machine Learning Tools and Techniques*. *Journal of Chemical Information and Modeling*. DOI: 10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.
12. Wheeler, D., & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*. DOI: 10.1007/s10109-005-0155-6.
13. Wheeler, D.C., Páez, A., Spinney, J., & Waller, L.A. (2014). A Bayesian approach to hedonic price analysis. *Papers in Regional Science*. DOI: 10.1111/pirs.12003.
14. Yeh, I.-C., & 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.
15. Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*. DOI: 10.1016/0165-0114(78)90029-5.