doi:10.20944/preprints201801.0262.v1

*Article*

# Index Cost Estimation Using Case Based Reasoning Model Based on Macro BIM

**Krzysztof Zima [1],*and Agnieszka Leśniak [2]**

[1] Institute of Construction Management, Cracow University of Technology, Warszawska 24 st., 31-155 Kraków, Poland; kzima@L3.pk.edu.pl
[2] Institute of Construction Management, Cracow University of Technology, Warszawska 24 st., 31-155 Kraków, Poland; alesniak@L3.pk.edu.pl
* Correspondence: kzima@L3.pk.edu.pl; Tel.: +48-12-628-2354

**Abstract:** Information regarding the cost of a construction project is available to the investor and project participants in order to determine the subsequent success of a project, given that the information they collect has an impact on the decisions they make. Cost calculations, especially in the initial phase of a project, often generate large errors. This paper presents the new approach based on a combination of the Case Based Reasoning method (CBR) with the originally selected criteria for the description of a construction project (as a result of Pearson correlation coefficient and Spearman's rank correlation coefficient) and Building Information Modeling (BIM) technology. The CBR method fulfils expectations for a simple and fast system supporting the cost estimation process. It does not require any specialist knowledge, so it will be comprehensible to cost estimation practitioners. The BIM-based model gives the opportunity for the calculation of quantity take-offs and enables the use of the information contained in the BIM model in the cost estimation process. In order to prepare the model an appropriate relational database had to be developed. With extensive research, a database of 173 construction projects, including the construction of a sports field, was obtained. There were 14 variables defined originally by authors; however, only 10 (as a result of the correlation analysis) were used for the calculation. Data related to the project were collected in the BIM model. Results estimating the project's unit price, using the CBR method, were presented and discussed. The Mean Absolute Estimate Error was used to evaluate the model.

**Keywords:** Building Information Modeling; Case Based Reasoning; cost estimating; information management

---

## 1. Introduction

The collection and storage of information is one of the most important tasks undertaken during a construction project. The impartial information received as project utility evaluation is useful for strategic planning, quality management, for solving the tasks of resource allocation, motivational project evaluation [1]. The information the investor and other project participants have access to determines the subsequent success of the project given that the information they collect has an impact on their decisions. For owners, contractors, and other stakeholders, forecasting the construction cost trend is very important to accurately estimate construction costs, prepare the project budget, control costs, and assess the associated risks [2]. As wrote Kapliński and Tupenaite one of the greatest challenges in the modern construction economics is efficient decision-making [3].

Cost is one of the major criteria in decision making at the early stages of a building design process [4]. The decision about the continuation of investment is mainly based on the cost of the building project. Performance and overall project success are often measured by how well the actual cost compares to the early cost estimates [5].

Cost estimate accuracy research has been the basis for the creation of a model for supporting cost estimation in the early phase of the project. The need for a correct and possibly accurate cost estimate in this phase is very important to the investor and impacts the decision-making on the continuation of the project and its subsequent success.

The literature contains many models supporting cost estimation of construction works. Despite the fact that there are many different prediction models, the improvement of prediction accuracy is still an acute problem that is facing decision makers in many areas [6]. These models can be divided according to two criteria: the type of the cost estimate and the mathematical method used in the construction project cost estimation model. Two kinds of estimates can be distinguished within the cost estimate type criterion: the initial estimates performed at the concept stage and the detailed estimates performed at the contractor selection stage. In turn, models supporting cost estimation can be divided according to the mathematical methods used as follows:

- Models using artificial neural networks,
- Models based on fuzzy set theory,
- Models based on analytic hierarchy process (AHP),
- Models based on multiple regression,
- Models based on case-based reasoning (CBR),
- Hybrid models using at least two different mathematical methods.

There are many models supporting cost estimation at the initial cost estimation stage in the literature. For example, Lowe, Emsley and Harding [7] used linear regression models to predict construction costs. The authors based their research on data from 286 construction projects implemented in the United Kingdom. Six models are developed for cost/m2, log of cost, and log of cost/m2. Forty-one potential independent variables were identified by authors and five variables appeared in each of the six models: gross internal floor area GIFA, function, duration, mechanical installations, and piling, suggesting that they are the key linear cost drivers in the data.

Gunaydin and Dogan in [4] and Juszczyk in [8] used artificial neural networks for cost estimation in the initial project phase. Gunaydin and Dogan estimated the costs of 4-8 floor apartment buildings in the design phase, erected using reinforced concrete technology. In order to forecast the costs, they used 8 criteria: total area of the building, ratio of the typical floor area to the total area of the building, ratio of ground floor area to the total area of the building, number of floors, console direction of the building, foundation system of the building, floor type of the building and location of the core of the building. Juszczyk performed an analysis of a cost estimation documentation for 102 multi-family dwelling houses built between 2003 and 2008. The author's analyses used 13 explanatory variables such as: footprint area, floor spaces of the flats, commercial units and underground garages, cubic volume of the building, number of floors, type of foundation, building structure and roof structure, number of structural segments of the building and lift shafts, ground conditions and standard of finish. The author applied both index methods and single neural networks and groups of artificial neural networks to model the costs of the buildings.

The case-based reasoning was used, for example, by Ji, Park and Lee in [9] and Marzouk and Ahmed [10]. Ji, Park and Lee based their analysis on a database of actual costs of public projects in Korea. The projects concerned 164 apartment buildings (104 cases from 2005, 28 from 2007, and 32 from 2009). The authors used 12 criteria in their research: Number of households, Gross floor area, Number of unit floor households, Number of elevators, Number of floors, Number of piloti with household scale, Number of households of unit floor per elevator, Height between stories, Depth of pit, Roof type, Hallway type, Structure type.

Marzouk and Ahmed in their research [10] presents a parametric-cost model for pump station projects. Fourteen factors have been identified as important to the influence of the cost of pump station projects. A data set that consists of forty-four pump station projects (fifteen water and twenty-nine waste water) are collected to build a Case-Based Reasoning (CBR) library and to test its performance.

The cost estimation model using the fuzzy set theory was proposed and described by El Sawalhi [11]. The author conducted surveys and used the relative index ranking technique to identify 5 factors that have the greatest impact on the cost of a construction project. The most important factors were: typical floor surface area, number of floors, number of lifts, cubic volume occupied by the HVAC (Heating, Ventilation, Air Conditioning) systems and the type of exterior finish. Each analysed design (106 construction projects in Gaza Strip) provided a source of information and included a cost estimate and an actual budget and final reports. It is also possible to find cost estimation models that use hybrid methods in the literature.

For example, Cheng M-Y., Tsai H-C., Hsieh W-S [12] used an evolutionary fuzzy neural interference model for cost estimation and Cheng M-Y., Tsai H-C., Sudjono E. [13] used neural networks with fuzzy criteria evaluation. An, Kim and Kang [14] used the CBR reasoning with the determination of the validity of criteria via the AHP process and Kim and Shim [15] supported the CBR reasoning process with genetic algorithms.

BIM-based models supporting cost estimation are currently described in the literature mainly because of the opportunities provided by the BIM model for quantity take-offs, and conceivably due to the possibility of using the information contained in the BIM model in the cost estimation process. Actually designing is a team game, that, in order to cope with, you need a competent team at hand and new organizational methods should be used (e.g., integrated design), new tools (e.g., Building Information Modeling) [16]. Consequently, the opportunities and benefits of using the BIM model in the quantity take-off process were discussed, among others, by [17-18]. The cost estimation using traditional cost estimation formulas and based on the BIM model saved in the IFC format was discussed, among others, by [19-21].

The aim of the article is to present the method of cost calculation at the initial stage of a construction project intended for the investor, which meets two conditions: it is easy to apply and generates a relatively low calculation error in relation to the costs of bids. For this purpose, the authors have chosen the method of case based reasoning, which is based on the actual case database and, thanks to case similarity measurement, shows not only the cost adopted on the basis of the selected case, but also gives the case similarity value allowing to assess the solution's adjustment. The BIM model was used due to quick and easy of calculating the amount of work, and the ability to store data in an IFC file.

## 2. CBR and BIM in supporting the estimation of project costs

Materials and Like the rest of the EU, Poland is currently facing the prospect of implementing and applying the EU Directive on the application of the BIM concept in public procurement contracts. It should be assumed that public and private investors will soon see the benefits of using the BIM technology in construction projects.

The advantage of using the BIM model is the considerable acceleration and automation of the quantity take-off process for construction works and their valuation. A take-off based on a building model saved in the IFC format makes it possible to read the amount of works in any form instead of laboriously calculating the amount of works to be performed. Taking advantage of the possibilities provided by the BIM model for the geometry of construction elements allows for quick and accurate calculations.

The accessibility of the information stored in one place in the model and the virtual representation of the planned civil structure in 3D make it possible to thoroughly analyse the individual building elements.

Recognising BIM as a dynamically developing technology and appreciating its numerous advantages for the take-off and costing processes, the author adopted the macro BIM model as a basis for the model supporting cost estimation. The macro BIM model is a building information model with a low level of detail, used in the phase of preparing the construction and investment process and applied for a macro-level analysis, including visualisation, spatial modelling, cost calculation, etc. The macro model allows for index calculations based on information about the size of the civil structure, such as its cubic volume, footprint area, floor space, number of floors, etc.

The choice of the CBR reasoning method was dictated by its continuous learning mechanism independent of the expert and the low requirements for acquired knowledge or the lack of precision in the possessed information and comparatively simple and fast calculation. An additional advantage of using the CBR method is that it relies on a database containing historic price data from previously completed projects instead of prices from price bulletins.

For the CBR method, a rule-based specialist knowledge is not required as is the case with most expert systems. A disadvantage of these methods is the difficulty in formulating an appropriate set of rules and often a complex method of its verification and supplementation. The cost estimation process is characterised by certain dynamics of changes in the environment, understood as all the changes affecting the prices of the construction works. However, the construction market cannot be considered as a fully stable environment. When analysing methods for the construction of intelligent systems, it is clear that the CBR method is best-suited to meet the adopted assumptions, if the assumptions about the environment and the required knowledge are taken into consideration.

The CBR method, therefore, fulfils expectations for a simple and fast system supporting the cost estimation process. It does not require a specialist knowledge, so it will be understandable to cost estimation practitioners, which is an additional reason for using this method.

## 3. Relational database for cost estimation purposes

The collection and management of knowledge is the basis of any information-based system. For cost estimates to be realistic and precise, they should be based on information from previously completed projects. The actually incurred costs of the construction works and the costs resulting from the cost estimates provided in the bids are a priceless source of knowledge for cost calculations performed in the early project phase.

A database is a set of interrelated information that can be stored, for example, in a BIM model. On the other hand, a relational database can create a set of unordered tables that can be used in analysis or report generation by using an operation that returns entire tables in its result [22]. The purpose of creating a relational IFC-DB database based on the BIM model is to organise and systematise the data on the costs of construction works and building elements obtained from previous construction projects. The information must be prepared in a way that enables their use in initial estimations, so it should allow for index calculations. The IFC-DB database in question is based on the completed cost estimates provided in bids for civil structures. The IFC-DB database was created on the basis of bidding cost estimates, descriptions of procurements for construction works and design documentation. Each construction project was described in the database with the available information, such as the cost of the construction project, the size of the works, the date of completion, the location of the civil structure, etc. Figure 1 shows the concept of the database created on the basis of construction designs and cost estimate calculations.

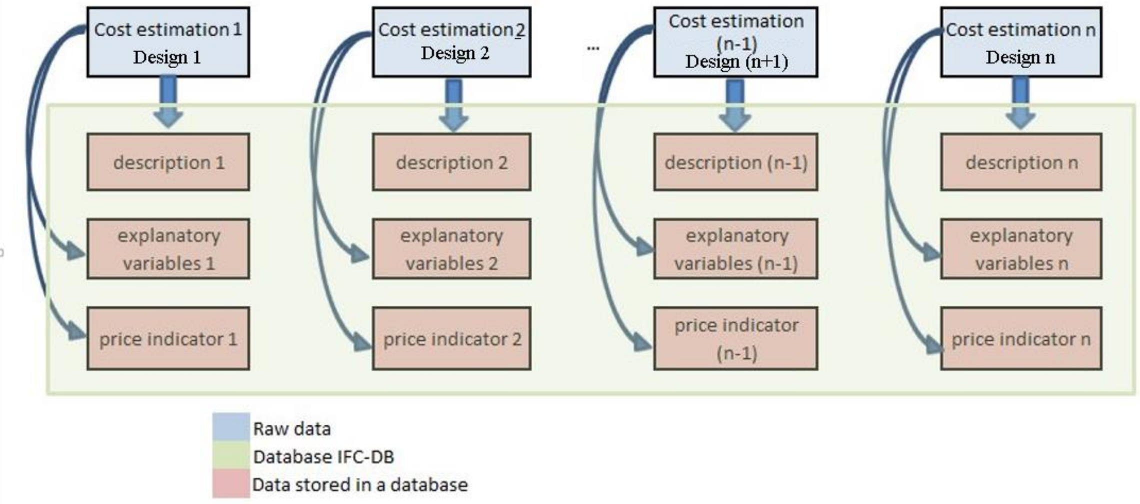


**Figure 1.** Creation of a database supporting the simplified cost estimation of construction works

The information included in the IFC-DB database can be obtained from BIM models of civil structures saved in the IFC format, especially in the area of the civil structure geometry, and from the cost estimates recorded together with the BIM model of the civil structure. This will allow for the use of the information collected in models, as proposed by the BIM concept. Such a database using BIM models has the added advantage of being able to easily store and retrieve geometric data and automatically retrieve data directly from the model saved in the IFC format. The IFC-DB database developed to support cost estimation with the CBR method includes old cases containing essentially four groups of information: information on the geometry of the civil structure - G, solution - S, description of the situation - D and explanation - E. In order to formalise the description of the old cases, the IFC-DB database was defined by the following formula:

$$\mathrm{IFC-DB} = \bigcup_{\mathrm{i=1}}^{\mathrm{n}} Case_i\{G_i, \mathrm{S_i}, \mathrm{D_i}, \mathrm{E_i}\} \qquad (1)$$

where:

$Case_i$ {$G_i$, $S_i$, $Di$, $E_i$} - i-th case extracted from the BIM model,

$G_i$ - graphical representation of the model elements for the i-th case,

$S_i$ - solution, i.e. the costs of construction of the elements included in the BIM model,

$D_i$ - cost estimate situation for the i-th case (time of completion, place),

$E_i$ - description of the construction project for the i-th costed case,

n - number of old cases in the database.

Solution $S_i$ contains price value indices in any monetary unit referenced to surface area or cubic volume measurement units as well as a description of the situation $D_i$, which contains the necessary information describing the market factors affecting the cost of the project and the description of the construction project $E_i$ in the form of qualitative data describing the civil structure.

In the case of the simplified cost estimation in the initial phase of the project, the graphical representation of the model elements is limited to surface and cubic volume indices. The solution is the total cost of construction of the civil structure or a unit cost. The cost estimate situation includes the parameters used to adapt the solution, including data on the location of the project and the date of completion. The case description includes the information on the civil structure that affects the cost of its construction to a specified degree. The explanatory variables describing the case may vary depending on the type of civil structure. The limit of the impact on the cost of the project, which constitutes the criterion for the selection of the final set of explanatory variables, is determined by the decision maker based on, for example, correlation analysis between the variables describing the project and the total or unit cost of construction of the civil structure.

## 4. Supporting the cost calculations based on the Macro BIM model

The paper presents the ICE-MACRO (Index Cost Estimate-MACRO) method supporting the calculation of costs, which was created by the author to determine the index cost estimates based on the Macro BIM model. The method is dedicated to the 2nd or 3rd level of maturity of the BIM model. The level of development of the BIM model in accordance with the AIA (American Institute of Architects) classification is LOD 100 or LOD 200, i.e. it is the level of development of the model at the concept or conceptual design stage. A decision maker or cost estimator performing the cost estimation of the construction project at the early phase of the project has only the general parameters of the planned civil structure. The ICE-MACRO method is based on the relational IFC-DB database discussed earlier and the model of the civil structure saved in the IFC format that represents the new case. Figure 2 shows the procedural algorithm used in the ICE-MACRO method.

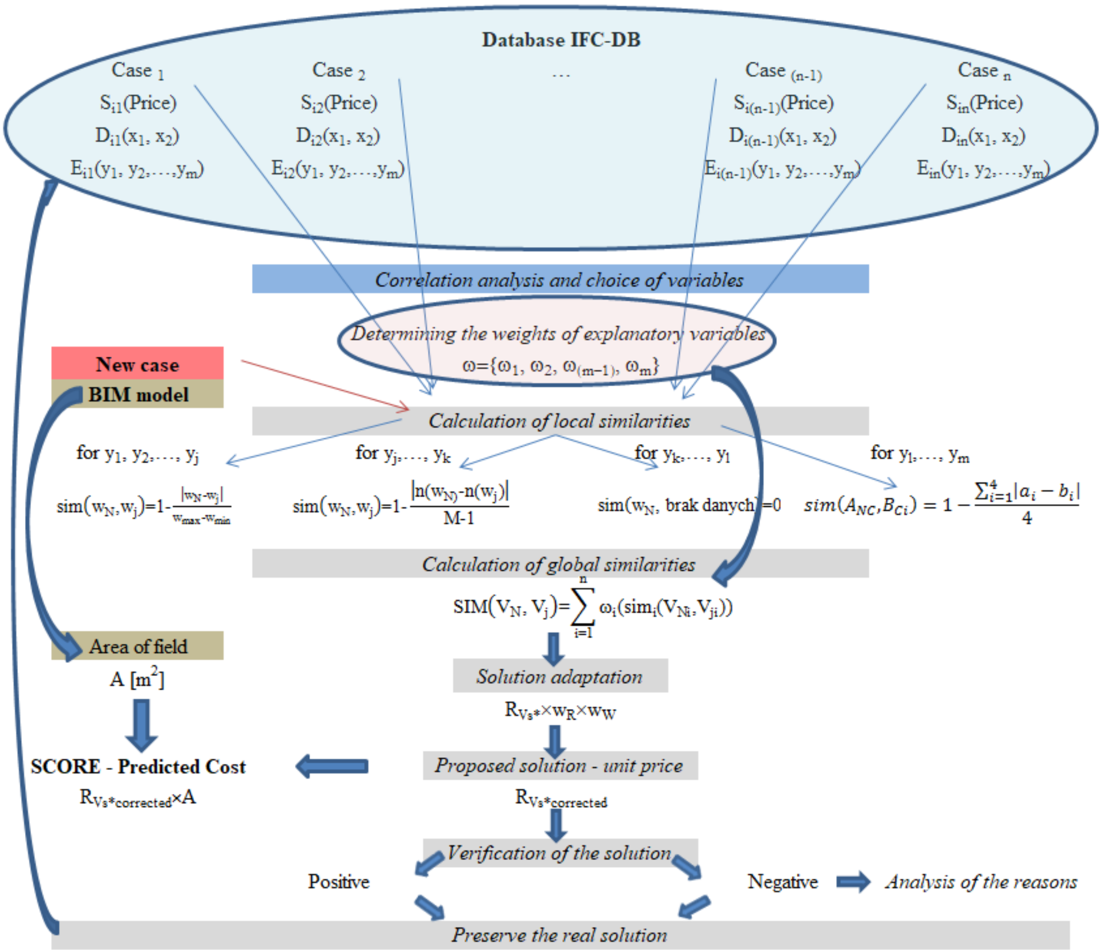


**Figure 2.** Procedural algorithm used in the ICE-MACRO method

The problem to be solved is the forecasted cost of constructing the civil structure represented by the BIM model. Currently, BIM models are most commonly saved in the IFC format, enabling the exchange of information with many project participants who use different software. The IFC format is currently the most popular common data scheme for maintaining and exchanging data between different BIM applications used in the construction industry. It is a neutral and open specification, which is the basic data file format based on the data model developed by the buildingSMART alliance® in order to facilitate interoperability in the construction industry. This format improves communication, reduces the time of delivery of the civil structure data and increases productivity and quality throughout the building's lifecycle.

The information that can be contained in a model saved in accordance with the IFC specification includes:

- Hierarchical information about the building (phase, stage of construction of the civil structure),
- Information about the type of an element (walls, ceilings, columns, beams, stairs, etc.),
- Information about the geometry of the civil structure and its elements (dimensions, coordinates of the element, surfaces and volume),
- Information about relationships between individual elements (e.g. connections between elements),
- Information about standard and non-standard properties assigned to elements (material, colour, cross-sections, fire protection, weight, etc.),
- Information on the cost of construction of elements, time and required quality.

The first step in the presented method is to specify the explanatory variables $E_i$ affecting the cost of the civil structure. There can be any number of variables, which varies depending on the type of the civil structure. The explanatory variables should be preferably strongly correlated with the response variable, i.e. the cost of the construction of the civil structure, and poorly correlated with each other. The next step is to specify the validity of individual explanatory variables. The validity of the variables will depend on the degree of impact on the cost of the civil structure, i.e. the strength of correlation with the variable Cost. The correlation describes the strength and type of relationship between two variables, with the strength of the relationship being described by a number and the type being described by a '+' or '-' sign. The correlation calculations used the Pearson correlation coefficient and Spearman's rank correlation coefficient. The Pearson coefficient measures the strength of the linear relationship between variables, where the variables are of a quantitative nature, and Spearman's rank correlation coefficient measures monotonic relationships. Therefore, Pearson correlation coefficient was used only to examine the strength and direction of the relationship between the measurable variables, and Spearman's rank correlation coefficient was used to describe the strength of correlation of two variables where the examined variables were of a qualitative nature.

The next step is the introduction of an appropriately chosen computational mechanism used in the application of the CBR method that is aimed at determining the degree of similarity between the description and the currently analysed new case, using the case-specific explanatory variables found in the IFC-DB database. Thanks to the used algorithms, the system user can obtain both information and solutions at the level of the structure of the member function and the solutions for the completion of the overall problem [23]

Therefore, all the old cases contained in the IFC-DB database should be compared with the new case according to the CBR method that measures the similarity of new and old cases. Similarity measurements are performed according to the nature of the explanatory variable.

The similarity for quantitative variables or ordinal variables determining the amount of construction works to be performed, or the quantities that characterise the civil structure, was measured according to the formula (as per [24]):

$$\mathrm{sim}(w_N,w_j)=1-\frac{|w_N-w_j|}{w_{max}-w_{min}} \tag{2}$$

where:

$w_N$ - value of the explanatory variable for the new case,

$w_j$ - value of the explanatory variable for the j-th old case,

$w_{max}$, $w_{min}$ - minimum and maximum values for all the old cases included in the database.

The similarity for qualitative variables determining the amount of construction works to be performed or the quantities that characterise the civil structure was measured according to the formula:

$$\mathrm{sim}(w_N,w_j)=1-\frac{|n(w_N)-n(w_j)|}{M-1} \tag{3}$$

where:

$n(w_N)$, $n(w_j)$ - place in an ordered array of values $n(w)=1,2,\dots, n$,

M - number of values.

If one or both of the variable values are unknown, the similarity value equals zero:

$$\mathrm{sim}(w_N,\text{ no data})=0 \tag{4}$$

There may also be a situation where there is unclear or inaccurate data. In such a case, it is possible to use mathematical operations based on fuzzy logic. It is then necessary to record the inaccurate information about the civil structure or its elements included in the BIM model in the form of a fuzzy number and determine the shape of the membership function. The use of fuzzy logic will make it possible to use intermediate values from the set [0, 1] when evaluating the explanatory variables. The calculation formula depends, however, on the assumed shape of the membership function for the given explanatory variable. For example, taking into account two trapezoidal fuzzy numbers $A_{NC} = (a_1, a_2, a_3, a_4)$ and $B_{Ci} = (b_1, b_2, b_3, b_4)$, the similarity SIM($A_{NC}$, $B_{Ci}$) can be defined by the following formula:

$$sim(A_{NC}, B_{Ci}) = 1 - \frac{\sum_{i=1}^{4}|a_i - b_i|}{4} \quad (5)$$

where:
sim($A_{NC}$, $B_{Ci}$) - similarity between fuzzy numbers,
$A_{NC}$ - fuzzy number for the new case,
$B_{Ci}$ - fuzzy number for the old case taken from the database,
$a_i$, $b_i$ - characteristic points for fuzzy numbers $A_{NC}$ and $B_{Ci}$.

In the presented algorithm, it was decided to separately calculate local similarities using the formulas (2), (3), (4) and (5) above. After analysing the local similarities, i.e. the similarities generated for individual variables, the global similarity should be calculated using the weights assigned to individual variables:

$$\mathrm{SIM}(V_N, V_{Sj}) = \sum_{i=1}^{n} \omega_i(\mathrm{sim}_i(V_{Ni}, V_{Sji})) \quad (6)$$

where:
$\omega_i$ - weight of the i-th explanatory variable,
SIM($V_N$,$V_j$) - global similarity between the old Vj and the new case VN,
$\mathrm{sim}_i$($V_{Ni}$,$VS_{ji}$) - local similarity for the i-th explanatory variable between the old $V_j$ and the new case $V_N$.

The case found should meet the following conditions after the calculation of the similarities between the cases:

1. The cases that have the highest value of the global similarity SIM($V_N$, $V_j$) are selected;
2. The minimum preferred similarity was determined arbitrarily by the author at 70%;
3. The similarity is calculated in natural numbers;
4. The minimum number of selected old cases is three (the value adopted by analogy with the rules used in property valuation, as the minimum number of the most similar properties selected for calculation in property valuation in the pair comparison method). One case with the similarity of up to 100% may be specific and have an overestimated or underestimated unit price value. Three or more cases guarantee greater reliability;
5. Cases significantly deviating from the rest of the selected cases (difference greater than 50% from the value of the other cases) in the situation of a greater number of selected cases (more than 2) are rejected;
6. The resulting unit price for the new case is a weighted average of the selected old cases, where the weight depends on the calculated similarity of the cases.

The solutions that meet these conditions are selected for the next stage, which consists in the adaptation of solutions.

The ICE-MACRO method always performs adaptation due to the time difference between the calculation of the selected most similar old case $V_{S*}$ and the new case, as well as due to the location of the civil structures. The adjustment will be performed using the regional factor $w_R$ and the indexation factor $w_W$. The regional factor reflects the price differences occurring in different project locations between the new case $V_N$ and the old case $V_{S*}$. The indexation factor reflects the differences in the calculations caused by the passage of time between the old case $V_{S*}$ and the new present case $V_N$.

After the adaptation has been performed, the newly created knowledge constitutes the forecasted initial cost estimation of the construction of the civil structure. After having been used, this value is transferred to a quarantine where it awaits the practical verification of the analysis results. After the adjustment resulting from the practical verification, the new case is saved and added to the existing IFC-DB database.

## 5. Computational example

The computational example showing the application of the ICE-MACRO model is concerned with the construction of single- and multi-purpose sports fields. Information on the completed projects of this kind come from public announcements on the intention to conclude a contract for construction works in the period from 2014 to 2016. The database contains 173 construction projects.

The sports field database DBSF contains 4 groups of information: a graphical representation of the model elements - G, the solution, i.e. the costs of the construction of 1 m$^2$ of the sports field surface area - S, the cost estimate situation (date of announcement, location of the project) - D, and the explanatory variables - E.

The solution S is the unit price of m$^2$ of the sports field surface expressed in PLN/m$^2$. The explanatory variables include the information describing the civil structure, both the quantitative geometric information - G and the information that qualitatively describes the construction project - E. The initial set of explanatory variables resulting from the literature study, as well as the analysis of the availability of information in announcements on public procurements for the construction of sports fields (on such an early stage of project), include 14 variables:

1. Quantitative variables
- intended use of the field,
- surface area of the field,
- surface area of the access paths and routes,
- green surface area,
- surface area of the ball containment netting,
- fence length.

2. Qualitative variables
- type of the sports surface,
- type of the material for access routes,
- type of the fence,
- type of sports equipment - handball,
- type of sports equipment - volleyball,
- type of sports equipment - basketball,
- type of sports equipment - football,
- type of sports equipment - tennis.

The database DBSF containing 173 construction projects consisting in the construction of sports fields was saved in the form of the following records conveying the relevant previously mentioned information:

Case$_n$ {G$_i$ = (surface area of the field, surface area of the access paths and routes, green surface area, surface area of the ball containment netting, fence length); S$_i$ = (unit price of the field surface area); D$_i$ = (location, date of the bid); E$_i$ = (intended use of the field, sports surface type, material for the access paths and routes, type of the fence, type of the equipment in the 1/0 = YES/NO notation - handball-volleyball-basketball-football-tennis)},

The data in the example contained in the database DBSF does not come from the BIM models due to the low availability of such models in practice. The new case analysed in the example was implemented as a BIM model and will be analysed as such, which will simplify the calculation of the needed quantitative data.

Table 1 presents the correlation coefficients of individual variables with the unit price of m2 of the surface area of the sports field and weights of explanatory variables calculated on the basis of correlation coefficients. The correlation coefficient for variables 1-5 was calculated using the Pearson correlation coefficient, and the correlation coefficient for the other variables was calculated using Spearman's rank correlation coefficient.

**Table 1.** The correlation coefficients and the weight of individual explanatory variables

| Variables | Correlation coefficient: | Weights $\omega_i$ |
|---|---|---|
| Surface area of the fields | -0.188 | 6.1% |
| Surface area of the access paths | 0.223 | 7.2% |
| Green surface area | 0.099 | rejected |
| Fence length | 0.257 | 8.3% |
| Surface area of the ball containment netting | 0.079 | rejected |
| Intended use | -0.492 | 15.8% |
| Pavement type | -0.300 | 9.7% |
| Material for access paths | -0.279 | 9.0% |
| Handball | 0.247 | 8.0% |
| Basketball | 0.464 | 14.9% |
| Volleyball | 0.359 | 11.5% |
| Football | -0.296 | 9.5% |
| Tennis | -0.065 | rejected |
| Fence type | -0.012 | rejected |

The assessment of the correlation strength used the classification of the interrelationship strength according to J. Guilford:

$r = 0$ no correlation,
$0 < r < 0.1$ barely perceptible correlation,
$0.1 < r < 0.3$ poor correlation,
$0.3 < r < 0.5$ average correlation,
$0.5 < r < 0.7$ high correlation,
$0.7 < r < 0.9$ very high correlation,
$0.9 < r < 1$ almost full correlation,
$r = 1$ full correlation.

Based on the assumption that the cost estimation should be as accurate as possible, and also taking into consideration the fact that there are few variables in the case in question, a decision was made to reject the variables that had no correlation or had barely perceivable correlation (variables 3, 5, 13, 14) and to keep only those that are characterised by at least poor correlation. The weights of the variables $\omega_i$ were calculated as a relationship between the absolute value of a given correlation coefficient and the sum of the absolute values of all correlation coefficients after the previous rejection of variables 3, 5, 13 and 14.

12 test cases (10%) were randomly selected from a set of 120 cases for subsequent testing as follows: Case 11, Case 24, Case 34, Case 35, Case 41, Case 48, Case 52, Case 68, Case 70, Case 75, Case 111 and Case 116.

Local similarities and the global similarity were calculated subsequently for all new cases. The examples of the detailed calculations of the project price for test case 11 are shown below.

Local similarities were calculated for subsequent old cases from the database by using formulas (2) for the quantitative variables and (3) and (4) for the qualitative variables (Tab. 2). Then, the calculations of the global similarity were performed (Tab. 3) according to the formula (6), using the weights of the variables in Table 1. The highest similarity value of 98% to test case 1 was achieved by: Case 7 - SIM($V_{\text{Test case 1}}$, $V_{\text{Case 7}}$); Case 13 - SIM($V_{\text{Test case 1}}$, $V_{\text{Case 13}}$) and Case 93 - SIM($V_{\text{Test case 1}}$, $V_{\text{Case 93}}$).

**Table 2.** The values of the local similarities and the global similarities for the cases with the highest similarity to New Case 1

| Variables | Local similarities | | |
|---|---|---|---|
| | Case 7 | Case 13 | Case 93 |
| Surface area of the fields | 0,94 | 0,94 | 0,99 |
| Surface area of the access paths | 0,97 | 0,97 | 0,99 |
| Fence length | 0,86 | 0,79 | 0,80 |
| Intended use | 1,00 | 1,00 | 1,00 |
| Pavement type | 1,00 | 1,00 | 1,00 |
| Material for access paths | 1,00 | 1,00 | 1,00 |
| Handball | 1,00 | 1,00 | 1,00 |
| Basketball | 1,00 | 1,00 | 1,00 |
| Volleyball | 1,00 | 1,00 | 1,00 |
| Football | 1,00 | 1,00 | 1,00 |
| ***Global similarities*** | ***0,98*** | ***0,98*** | ***0,98*** |

Table 3 shows a brief characteristic of the old cases selected during the calculation of the global similarities and test case 1.

**Table 3.** Selected cases with the highest degree of similarity to New Case 1

| Case | Test case 1 | Case 7 | Case 13 | Case 93 |
|---|---|---|---|---|
| No. of bid according to the public procurement office | 55769 - 2014 | 40988-2014 | 59010-2014 | 35376-2015 |
| Total price | 380 118,38 PLN | 186 435,09 PLN | 567 800,35 PLN | 467 710,00 PLN |
| Unit price | 287,97 PLN | 304,08 PLN | 985,76 PLN | 316,02 PLN |
| Date of estimation | 13.05.2014 | 06.03.2014 | 15.04.2014 | 25.03.2015 |
| Territory | Silesian region | Masovian district | Łódź province | Wielkopolska province |

Case 13 was rejected due to a considerable difference in unit price compared to the other two selected cases. After the analysis of the public procurement, it is clear that the content of the procurement was the construction of a multi-purpose sports field with accompanying elements, such as lighting and septic tanks, which do not fall within the scope of works in other procurements. The other two selected cases were adapted for the location of the project and the date of the cost estimate. Regional coefficients calculated by the author on the basis of the Sekocenbud regional price bulletin [25] were used in order to adjust the location. The Sekocenbud forecasting and indexation bulletin [26] was used to update the prices of construction projects estimated in the past. The adjusted unit prices are shown in Table 4.

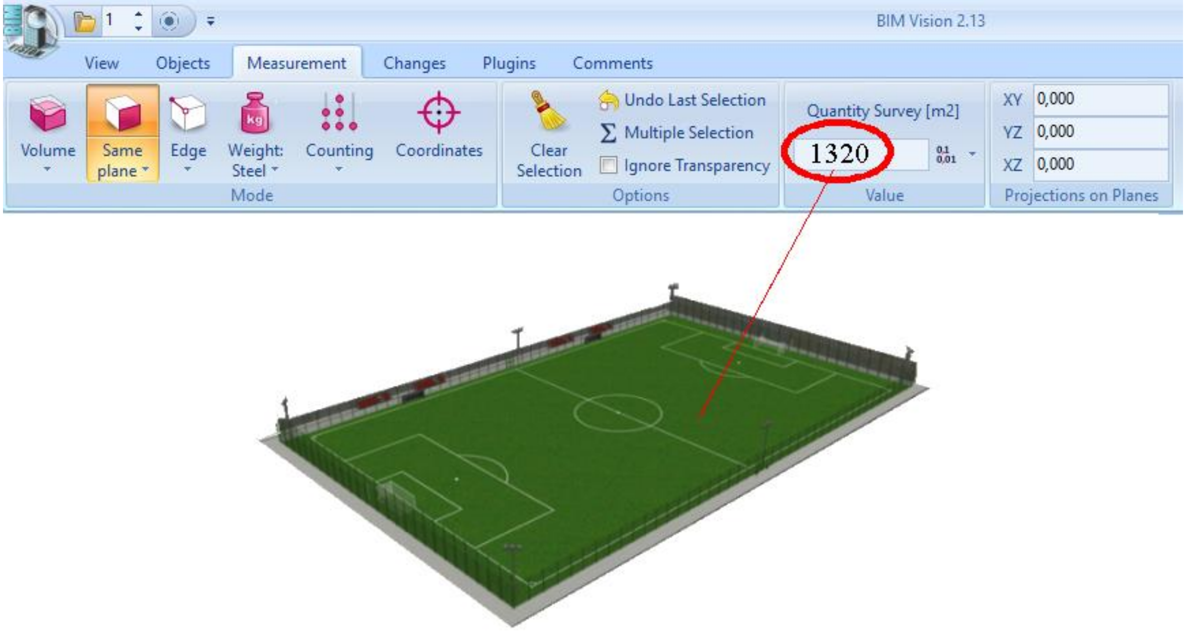
**Table 4.** Adjusted unit prices of two selected cases - case 7 and case 93

| Case | Case 7 | Case 93 |
|---|---|---|
| Unit price | 304,08 PLN | 316,02 PLN |
| Regional factor | 1,028 | 0,960 |
| Indexation factor | 100,1% | 101,3% |
| Adjusted price | 312,91 PLN | 307,31 PLN |

The adjusted price of Case 7 was calculated from the product PLN 304.08 × 1.028 × 1.001; where the regional factor is - 1.028 and the indexation factor is equal to (1 + 0.01%) = 100.1%. In Case 93, the adjusted price was calculated from the product PLN 316.02 × 0.960 × 1.013; where the regional factor is - 0.960 and the indexation factor is equal to (1 - 0.3%) × (1 - 0.0%) × (1 + 1.6%) = 101.3%. The final price for test case 1 is an arithmetic mean of the unit prices of case 7 and case 93 and equals:

$$\frac{(C_{jA}^{Case\ 7}+C_{jA}^{Case\ 93})}{2}=\frac{(312{,}91+307{,}31)}{2}=310{,}11\ PLN/m^2 \tag{7}$$

After calculating the price of test case 1, the surface area of the field was calculated from the BIM model of the civil structure. The surface calculations can be done very quickly and simply by selecting the surface in the sports field model view. An example of such a procedure is shown in Figure 3.



Quantity Survey [m2] 1320

XY 0,000

YZ 0,000

XZ 0,000

**Figure 3.** Take-off procedure in the ICE-MACRO method based on the MACRO BIM model - view in the BIM Vision browser

The final forecasted cost of the construction of the sports field for test case 1 is: 310.11 [PLN/m$^2$] ×1320 [m$^2$] = 409,340.67 PLN.

## 4. Discussion

Table 5 shows the results of the forecast of the unit price using the CBR method. Actual unit prices for m$^2$ of sports field surface taken from the winning bids, selected during the public procurement procedure for the test set (test cases 1-12) and the forecasted prices calculated using the ICE-MACRO method, were provided. The assessment was based on the calculated Mean Absolute Estimate Error (MAEE):

$$MAEE=\frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t-F_t}{A_t}\right| \tag{8}$$

where:
$A_t$ - current actual price,
$F_t$ - forecasted price,
n - size of the test set.

The error is smaller than 10% for test cases 1, 2, 6, 9 and 10 and the error exceeds the value of 20% only in cases 5, 8 and 11. Probably the reason for such large differences may be a relatively small database as well as the differences in the planned scope of the construction works and the difficulty in taking into account the scope of some works, such as demolition works at the initial stage of the project.

**Table 5.** The results of unit price estimation using the ICE-MACRO method together with the MAEE error

| Case | Real unit price of New case | Unit price calculated for New case | MAEE |
|---|---|---|---|
| 1 | 287,97 PLN | 310,11 PLN | 7,7% |
| 2 | 199,93 PLN | 195,72 PLN | 2,1% |
| 3 | 97,66 PLN | 82,10 PLN | 15,9% |
| 4 | 263,15 PLN | 212,43 PLN | 19,3% |
| 5 | 408,81 PLN | 323,57 PLN | 20,9% |
| 6 | 335,83 PLN | 320,97 PLN | 4,4% |
| 7 | 613,68 PLN | 527,82 PLN | 14,0% |
| 8 | 395,88 PLN | 312,75 PLN | 21,0% |
| 9 | 541,26 PLN | 491,19 PLN | 9,3% |
| 10 | 412,82 PLN | 437,88 PLN | 6,1% |
| 11 | 599,26 PLN | 407,46 PLN | 32,0% |
| 12 | 366,59 PLN | 318,54 PLN | 13,1% |
| **Total error** | | | **13,8%** |

The total MAEE error of 13.8% is acceptable, assuming that the accuracy range of the initial estimates is from -30% to +50% according to PMI 2008 [27]. The requirements of the American Association of Cost Engineers (AACE) - the error range of -10% to +15% - are also met. In addition, the estimated error range of -30% to +50% meets the requirements of the Construction Industry Institute (for: Kim and Shim 2014). It is worth noting that the error is significantly smaller than the error resulting from the actual differences between the investor's calculation and the winning bid. In Poland, the differences between the price of the selected bid and the value of the procurement resulting from the investor's cost estimate are very large. The maximum difference for the construction of sports facilities in the period between 2014 and 2015 was as much as 83.3%. The most frequent difference between the price of the selected bid and the estimated value of the procurement contract was 20% - 50%.

By comparing the results of the ICE-MACRO model with the results of the authors cited at the beginning of the paper, it should be noted that the results are promising. The ICE-MACRO method produces better results than the best model using multiple regression techniques described by Lowe et al. [7] and the index methods, which generate an even higher error rate.

The neural networks described by Gunaydin and Dogan in [4] and Juszczyk in [8] show a similar error rate for individual networks. Only neural network groups exhibit a large potential, generating a mean error of 6.04%. The problem with neural networks is the maximum percentage error $PE_{max}$, which is over 50% for both methods, as compared to the 32% error of the ICE-MACRO method, as well as the need for testing, learning and validating the networks, which generates the need to build a very large database. The advantage of the ICE-MACRO model is that it is also based on the BIM model, which enables the collection of information about the civil structure as well as fast and easy access to data, such as take-off information.

## 5. Conclusions

Based on the analysis above, it can be stated that the ICE-MACRO method accomplishes its task. The errors occurring during the analysis are acceptable in the light of the requirements of various associations or publications. The ICE-MACRO model performs well in comparison to other models when it comes to the mean absolute estimate error and, in particular, the maximum percentage error. It should be noted that two mathematical methods - neural networks and case-based reasoning - prove useful in supporting cost estimation in the early phases of a construction project. The BIM model, in turn, allows for the storage of information about the construction project and makes it possible to quickly generate basic surface or cubic volume indices.

**Author Contributions:** The individual contribution and responsibilities of the authors were as follows:

Krzysztof Zima designed the research main idea and collected the data. Authors togeather analyzed the data and the obtained results. Agnieszka Leśniak provided extensive advice throughout the study results and methodology. All the authors have read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Zavadskas, E.K.; Vilutiene, T.; Turskis, Z.; Saparauskas, J. Multi-criteria analysis of Projects' performance in construction. *Archives of Civil and Mechanical Engineering* **2014**, *14*, 114–121.
2. Xu, J-W.; Moon, S. Stochastic Forecast of Construction Cost Index Using a Cointegrated Vector Autoregression Model. *Journal of Management in Engineering* **2013**, *29 (1)*, 10-18.
3. Kaplinski, O.; Tupenaite, L. Review of the multiple criteria decision making methods, intelligent and biometric systems applied in modern construction economics. *Transformations in Business & Economics* **2011**, *10*, 166-181.
4. Gunaydin, H.; Dogan, S. A neural network approach for early cost estimation of structural systems of buildings. *International Journal of Project Management* **2004**, *22 (7)*, 595–602.
5. Oberlender, G.D.; Trost, S.M. Predicting accuracy of early cost estimates based on estimate quality. *Journal of Construction Engineering and Management* **2001**, *127(3)*, 173-182.
6. Yazdani-Chamzini, A.; Zavadskas, E.K.; Antucheviciene, J.; Bausys, R. A Model for Shovel Capital Cost Estimation, Using a Hybrid Model of Multivariate Regression and Neural Networks. *Symmetry* **2017**, *9(12)*, 1-14. doi:10.3390/sym9120298
7. Lowe, D.; Emsley, M.; Harding, A. Predicting Construction Cost Using Multiple Regression Techniques. *Journal of Construction Engineering and Management* **2006**, *132 (7)*, 750–758.
8. Juszczyk, M. Application of committees of neural networks for conceptual cost estimation of residential buildings. *AIP Conf. Proc.* **2015**, 1648, 600008.
9. Ji, S-H.; Park, M.; Lee, H-S. Cost estimation model for building projects using case-based reasoning. *Canadian Journal of Civil Engineering* **2011**, *38(5)*, 570–581.
10. Marzouk, M.; Ahmed, R. A case-based reasoning approach for estimating the costs of pump station projects. *Journal of Advanced Research* **2011**, *2 (4)*, 289–295.
11. Sawalhi El, N.I. Modeling the Parametric Construction Project Cost Estimate using Fuzzy Logic. *International Journal of Emerging Technology and Advanced Engineering* **2012**, *2(4)*, 631-636.
12. Cheng, M-Y.; Tsai, H-C.; Hsieh, W-S. Web-based conceptual cost estimates for construction projects using Evolutionary Fuzzy Neural Inference Model. *Automation in Construction* **2009**, *18 (2)*, 164–172.
13. Cheng, M-Y.; Tsai, H-C.; Sudjono, E. Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. *Expert Systems with Applications* **2010**, *37 (6)*, 4224–4231.
14. An, S-H.; Kim, G-H.; Kang, K-I. A case-based reasoning cost estimating model using experience by analytic hierarchy process. *Building and Environment* **2007**, *42 (7)*, 2573–2579.
15. Kim, S.; Shim, J.H. Combining case-based reasoning with genetic algorithm optimization for preliminary cost estimation in construction industry. *Canadian Journal of Civil Engineering* **2014**, *41(1)*, 65-73.
16. Bonenberg, W.; Kapliński, O. The Architect and the Paradigms of Sustainable Development: A Review of Dilemmas. *Sustainability* **2018**, *10*, 100.
17. Kim, S-A.; Chin, S.; Yoon, S-W.; Shin, T-H.; Kim, Y-S.; Choi, C. Automated Building Information Modeling System for Building Interior to Improve Productivity of BIM-based Quantity Take-Off. Proceedings of International Symposium on Automation and Robotics in Construction (ISARC 2009), 2009, 492-496.
18. Monteiro, A.; Martins, J.P. A survey on modeling guidelines for quantity takeoff-oriented BIM-based design. *Automation in Construction* **2013**, *35 (11)*, 238–253.
19. Cheung, F.; Rihan, J.; Tah J.; Duce, D.; Kurul, E. Early stage multi-level cost estimation for schematic BIM models. *Automation in Construction* **2012**, *27 (11)*, 67–77.

20. Liu, H.; Lua, M.; Al-Hussein, M. BIM-based Integrated Framework for Detailed Cost Estimation and Schedule Planning of Construction Projects. Proceedings of The 31st International Symposium on Automation and Robotics in Construction and Mining (ISARC 2014), 286-294.
21. Ma, Z.; Wei, Z.; Zhang, X. Semi-automatic and specification-compliant cost estimation for tendering of building projects based on IFC data of design model. *Automation in Construction* **2013**, *30 (3)*, 126–135.
22. Zima, K. Relacyjna baza danych wspomagająca kosztorysowanie na podstawie wnioskowania z przypadków / Relational database supporting cost estimation based on case based reasoning. *Materiały Budowlane* **2016**, *6*, 109-111.
23. Ociepka, P. Zastosowanie metody CBR do wspomagania procesu projektowo-konstrukcyjnego / Application of CBR method to support design and construction process. *Wybrane problemy inżynierskie* **2011**, *2*, 287-291.
24. Traczyk, W. Inżynieria wiedzy / Knowledge engineering. Akademicka Oficyna Wydawnicza EXIT, Warszawa, Poland, 2010.
25. Sekocenbud. Regional price bulletin – BCR, 2014-2017.
26. Sekocenbud. Forecasting and indexation bulletin – ZWW, 2014-2017.
27. PMI. A guide to the project management body of knowledge. Project Management Institute. Project Management Process Groups Interactions and Process Poster, 2008.