

Forecasting of steel consumption with use of nearest neighbors method

Michał Rogalewicz^{1,*}, *Robert Sika*², *Paweł Popielarski*², and *Grzegorz Wytyk*³

¹Poznan University of Technology, Chair of Management and Production, Piotrowo 3 Street, Poland

²Poznan University of Technology, Institute of Material Technology, Piotrowo 3 Street, Poland

³BiB Systems LLC, Lutycka 11, Poland

Abstract. In the process of building a steel construction, its design is usually commissioned to the design office. Then a quotation is made and the finished offer is delivered to the customer. Its final shape is influenced by steel consumption to a great extent. Correct determination of the potential consumption of this material most often determines the profitability of the project. Because of a long waiting time for a final project from the design office, it is worthwhile to pre-analyze the project's profitability and feasibility using historical data on already realized orders. The paper presents an innovative approach to decision-making support in one of the Polish construction companies. The authors have defined and prioritized the most important factors that differentiate the executed orders and have the greatest impact on steel consumption. These are, among others: height and width of steel structure, number of aisles, type of roof, etc. Then they applied and adapted the method of k-nearest neighbors to the specificity of the discussed problem. The goal was to search a set of historical orders and find the most similar to the analyzed one. On this basis, consumption of steel can be estimated. The method was programmed within the EXPLOR application.

1 Introduction

Development of trade, globalization, as well as diversified needs of clients and development of modern production and computer technologies caused that the modern manufacturing and service companies compete in a global market, where client is the most important. That is why it is important to ensure high quality of products and services, suited to his needs and expectations, as well as supplying them in an acceptable time. In this point of view, improvement of processes inside a company [1,2,3], limitation of wastage [4,5], shortening of a production cycle [6] and changeover times, minimization of distribution costs and improvement of quality of offered services have become the primary goals of organizations. Realization of these goals should be performed in connection with application of newest achievements in science and technology – new technologies on one hand, but also methods of decision-making on the other hand [7,8,9].

* Corresponding author: michal.rogalewicz@put.poznan.pl

Large amounts of data gathered in big databases in scope of a daily functioning of a company are today not always appropriately used for building knowledge about processes and products. Wasting such a detailed source of information, both about a company itself and client requirements, preferences, habits and needs, is a fundamental mistake made in certain organizations. It seems that in the era of Industry 4.0 proper acquisition, processing, analysis and archiving of data gain importance and make for a faster reaction to dynamically changing clients' requirements, shorter time of introduction of a product into market, better estimation of costs of undertakings, as well as obtaining higher client satisfaction [10,11].

To face these challenges, certain methods can be used, allowing analysis of large datasets and/or automatic obtaining of rules and patterns, which are not obvious and trivial for a database user – the so-called Data Mining methods.

This paper presents application of one of the Data Mining methods – the k -nearest neighbors method – for a quick, preliminary estimation of steel consumption for building a construction ordered by a client. It facilitates presentation of a preliminary offer to a client, without a need of preparing a detailed project, which could take even several days. Assuming that a company receives several inquiries daily and it orders preparation of a detailed project to a design office, it is important to preliminarily estimate the costs and present them to a client. Only after initial acceptance, actual, detailed project is prepared. The studies were realized in one of Polish construction companies.

2 The k -nearest neighbors method

The Data Mining methods fulfill several functions. They enable **describing data/objects**, **classification of objects** (meaning assigning them to classes of a dependent variable, on the basis of values of independent variables describing a given object), **grouping of objects** (creating groups of objects the most similar to each other, on the basis of values of independent variables), **search for associations** (finding interdependencies between data, they are based upon co-appearance of values of particular variables) and **realization of the regression task** (search for a model of relation between independent variables and a dependent variable in a numerical form) [12,13].

An example of a method that allows realization of both classification and regression tasks is the k -nearest neighbors method. This paper focuses on using this method for a preliminary estimation of steel consumption for an inquiry of quotation sent by a client, so the regression task is the one being realized. The k -nearest neighbors method focuses on finding a determined number of objects located in the closest proximity to a new object and calculating a value of an unknown dependent variable on their basis (usually it is simply an average of a dependent variable value for the selected k objects). The distance between objects in a multidimensional space, for the quantitative variables, is defined using one of the several selected metrics, e.g. Euclidean, Manhattan or Chebyshev. The first one is the most commonly used – it represents a standard way of perceiving distance by humans in real world. For the qualitative variables, the “different from” function is used, taking the value of 0 when two objects have the same value of a feature and 1 if the value is different [13,14,15].

The k -nearest neighbors method requires normalization of variables, as it is possible that a variable taking high values will totally dominate the distance metric and influence of the other variables. It is also worth ensuring, that there will be no situation, in which many less important variables, for which the studied objects will have a very distant value, will dominate very relevant variables, for which this distance will be very small. To achieve this, the so-called axis stretching in a multidimensional space can be applied, meaning choosing appropriate weights for the more important independent variables [13].

The further part of this paper presents an example of use of the k -nearest neighbors method for estimation of steel consumption for the quotation inquiries coming from clients. This estimation is conducted on the basis of already realized, similar projects, archived in a database. The consumption is determined by an a priori defined index. In scope of the performed work, a dedicated software tool EXPLOR was designed. It realizes the following functions: 1C, MultiC and k -neighbors. Their development is presented in the further part, on specific examples.

3 Application of the k -nearest neighbors method for estimation of steel consumption

3.1 Assumptions for the used method and application

Aim of the work conducted by authors of this paper was development of a method of supporting the decision-making in scope of steel consumption estimation, for building a construction, which is a result of initial discussions with a client. The following assumptions were taken, regarding the used method and the designed software application:

- the application uses a database of projects already realized by the construction company – at the moment of launching the application, the database had approx. 300 records (where a single record contained data from one, specific order realized for a client),
- the most important features, differentiating the projects and having the most important influence on a level of steel consumption are:
 - hall width [m],
 - hall height, from ground to bottom of roof construction [m],
 - number of aisles [1 or 2],
 - roof type [penthouse or gable roof],
 - roof load [kN/m²],
- the records regarding halls with a gantry and without it,
- in the case of the k -neighbor function, the user was given a choice of number of objects (“neighbors”), on basis of which the estimation is conducted, the Euclidean distance was used,
- the application realizes three tasks, which are presented in a greater detail in chapter 3.2.

3.2 Exemplary use of the application

The authors prepared a software application, realizing three tasks aimed at adjusting value of an independent variable from the already realized orders, to a new offer, on the basis of given values of independent variables. The first two described tasks (1C, MultiC) mostly pertain to rapid comparison of new offers to already realized projects, while the k -neighbor method uses a selection of weights of independent values by the authors. Panel of the EXPLOR software is shown in Figure 1 and Figure 2.

The first function of the EXPLOR program (named the 1C) regards use of only one criterion of result adjustment – a specific value of an independent variable, e.g. finding records of an applied value of hall width. This function is particularly useful, when a company must analyze types of already realized projects by a specific, single criterion. The algorithm will often return a large amount of results (in comparison to the other two functions), which may be then used for deepened exploration, e.g. indicate the most profitable orders of the ones selected.

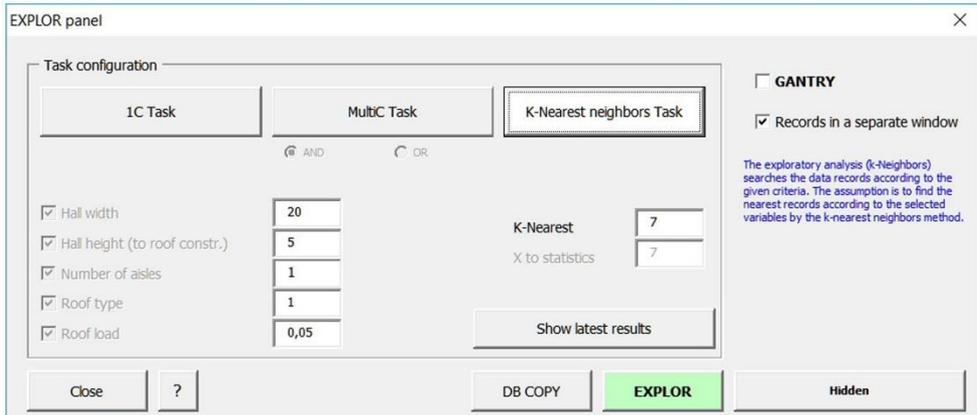


Fig. 1. Main panel of EXPLOR program

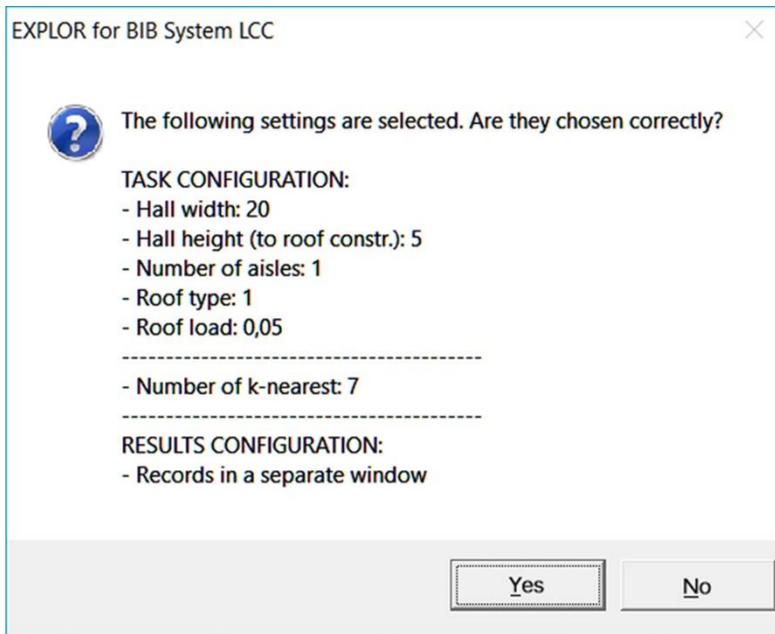


Fig. 2. Settings confirmation in the k-neighbor Task

The second function (named the MultiC) is an expansion to the 1C function. In this case, more than one criterion can be indicated (up to all possible) and logical operators AND and OR can be used. In the first case, results will be returned if there will be records fulfilling all the criteria for the input values. For example, for the following three criteria: hall width (20 m), hall height to bottom of the roof construction (5 m) and amount of aisles (1 aisle), the found records must all contain data about orders concerning 20-meter halls, height equal to 5 meters measured to roof bottom and having only 1 aisle. The second option – OR – returns results when at least one of the indicated criteria is fulfilled.

The last function – *k*-nearest – is the most expanded one. The idea is similar to the MultiC function with the AND operator (the found records will fulfill all the selected criteria for specific values of independent variables), but here the algorithm adjusts values indicated by a user to the closest contained in a database record, using data normalization and axis stretching in a multidimensional space. It is the most useful option in a situation of

estimating the steel consumption for quotation inquiries coming from clients. On the basis of experience of the company and the developed know-how, appropriate values of weights were selected for the particular independent variables. The greatest weight is assigned to the most important criterion. Figure 1 shows exemplary settings in the k-nearest function window, while in Figure 2 the settings confirmation in the k-neighbor Task was presented.

As presented, the steel consumption is determined on the basis of 7 nearest neighbors of the case analyzed (“new offer” – see Figure 3). The results will be returned in a separate window and the nearest neighbors are ordered starting from the smallest distance. It is noteworthy, that in this case, the results can be returned even if the found value of the independent variable is not exactly the same as indicated by a user. Figure 3 presents a results window. The analysis result is an average value of the steel consumption index, as well as the standard deviation.

	A	B	C	D	E	F	G	H	I
	Offer No.	Hall width [m]	Hall height to roof constr. [m]	Number of aisles	Roof type	Roof load [kN/m2]	Location	Steel consumption indicator [kg/m2]	
1									
2	new offer	20	5	1	1	0,05 ? ...	
3	1	20	5	1	2	0,05		28,00	
4	45	20	5	1	2	0,1		26,80	
5	186c	20	5	1	2	0,1		24,00	
6	186d	20	5	1	2	0,1		22,00	
7	220a	20	5	1	2	0,15		27,71	
8	47	20	5,5	1	2	0,05		26,67	
9	187d	20	5	1	2	0,25		25,00	
10								Xaverage	25,74
11								STD Deviation	2,18
12	Other results, less matching...								
13	46	20	6	1	2	0,1		26,00	
14	48	20	6	1	2	0,1		23,00	
15	235	24	3,8	1	1	0,15		26,49	
16	22	20	6	1	2	0,15		23,75	
17	219b	20	6	1	2	0,15		27,25	
18	182a	20	4	1	2	0,15		25,12	
19	148	17,5	6,5	1	1	0,1		31,89	
20	35b	13	3,8	1	1	0,1		19,66	
21	34	10	4,5	1	1	0,15		18,15	
22	4	27,05	5,535	2	1	0,15		32,60	

Fig. 3. Result of work of the k-nearest function. The result is 7 closest neighbors.

4 Conclusions

The paper presents an innovative approach to supporting the decision making in the Bidding and Contract Preparation department, on example of a construction company. The authors, together with employees of a company, defined factors having the highest influence on steel consumption and the most differentiating for the steel constructions and ordered them by importance. The possible ways of database exploration were presented, to find already realized orders, which are the most close to an analyzed order, on the basis of values of features defined by the user.

Facing the tight deadlines and difficulties with preparing even the preliminary projects for several or more than a dozen clients at the same time, it was decided to implement the solution proposed by the authors. Quick preparation of an offer using a steel consumption index estimated on the basis of data of already realized orders, allowed the company to perform calculations of profitability of undertakings in virtually no time, as well as to limit costs of project preparation.

The historical data, appropriately processed and analyzed, have turned out to be very useful in daily functioning of a company and building the process knowledge. The tool developed by the authors is universal and may be used by other companies, after modifying the most important independent variables and values of their weights.

References

1. A. Kujawińska, K. Vogt, *Man. and Prod. Eng. Rev.* **6**, 2 (2015)
2. A. Kujawińska, K. Vogt, A. Hamrol, *Adv. in Int. Sys. and Comp.*, Springer Verlag, **490** (2016)
3. A. Hamrol, *Prod. Plan. and Contr.*, **11**, 8 (2000)
4. M. R. L. Varela, J. Trojanowska, S. Carmo-Silva, N. M. L. Costa, J. Machado, *Man. and Prod. Eng. Rev.*, **8**, 2 (2017)
5. E. Dostatni, J. Diakun, A. Hamrol, W. Mazur, *Ind. Manag. and Data Sys.*, **113**, 6 (2013)
6. C. Gangala, M. Modi, V. K. Manupati, M. L. R. Varela, J. Machado, J. Trojanowska, In: A. Rocha, A. Correia, H. Adeli, L. Reis, S. Costanzo (eds) *Rec. Adv. in Information Sys. and Tech. WorldCIST 2017. Adv. in Int. Sys. and Comp.*, **571** (2017)
7. J. Trojanowska, M. L. R. Varela, J. Machado, In: A. Rocha, A. Correia, H. Adeli, L. Reis, S. Costanzo (eds) *Rec. Adv. in Information Sys. and Tech. WorldCIST 2017. Adv. in Int. Sys. and Comp.*, **571** (2017)
8. M. Diering, K. Dyczkowski, *IEEE Int. Conf. on Ind. Eng. and Eng. Man.* (2016)
9. A. Kujawińska, M. Rogalewicz, M. Diering, A. Hamrol, In: A. Rocha, A. Correia, H. Adeli, L. Reis, S. Costanzo (eds) *Rec. Adv. in Information Sys. and Tech. WorldCIST 2017. Adv. in Int. Sys. and Comp.*, **571** (2017)
10. Z. Ignaszak, R. Sika, M. Perzyk, A. Kochański, J. Kozłowski, *Arch. of Found. Eng.*, **145**, 5 (2016)
11. Gorski F., Zawadzki P., Hamrol A., *J. of Mach. Eng.*, **16**, 4 (2016)
12. J. Han, M. Kamber, J. Pei J., *Data Mining: Concepts and techniques*, (Morgan Kaufmann Series in Data Management Systems, Elsevier, 2012)
13. T. Larose, *Discovering knowledge in data: an introduction to Data Mining* (Wiley&sons, 2005)
14. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of statistical learning. Data Mining, inference, and prediction* (Springer, 2009)
15. T. Hill, P. Lewicki, *STATISTICS: methods and applications.* (StatSoft, Tulsa, 2007)