SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2023
ORGANIZATION AND MANAGEMENT SERIES NO. 184

APPLICATION OF ASSOCIATION RULES IN FILLING GAPS IN SURVEY DATA

Agnieszka RZEPKA[1], Adam KIERSZTYN[2], Radosław MIŚKIEWICZ[3*], Krystyna KIERSZTYN[4]

[1] Faculty of Management, Lublin University of Technology; a.rzepka@pollub.pl, ORCID: 0000-0003-4495-6066
[2] Faculty of Electrical Engineering and Computer Science, Lublin University of Technology; a.kiersztyn@pollub.pl, ORCID: 0000-0001-5222-8101
[3] Institute of Management, University of Szczecin; radoslaw.miskiewicz@usz.edu.pl, ORCID: 0000-0003-2370-4783
[4] Faculty of Natural Sciences and Technology, Catholic University of Lublin; krystyna.kiersztyn@.pl, ORCID: 0000-0003-1957-1797
* Correspondence author

Purpose: Surveys are one of the most popular data acquisition tools used in economics and management sciences. The results of surveys provide a lot of information and allow for fast response to changes in the socio-economic environment. Unfortunately, in many cases there are missing data in surveys, which can be caused by various reasons.
Design/methodology/approach: One of the most common reasons are the respondent's reluctance to provide an answer or distraction while completing the questionnaire. This study presents a novel approach for filling gaps in the survey data.
Findings: The main idea of the proposed method is to use the associations between the answers to given sets of questions for different respondents.
Originality/value: The obtained association rules were used as input variables and a number of well-known machine learning tools were applied for filling data gaps. The results of numerical experiments confirmed a very high performance of the proposed novel method for filling data gaps in surveys.

Keywords: Surveys, association rules, filling data gaps, Fuzzy Rule, Random Forest.

Category of the paper: Research paper.

1. Introduction

The diversity of research interests in the field of economic sciences (Baumans, Davis, 2010) is associated primarily with the breadth of research issues (Bennet, 2006; Burke, Morley, 2016). In social sciences, individual disciplines of economic sciences (Düppe, 2011; Farmer, 2013) have the characteristic which could be defined as multi-paradigmaticity (Friedman, 2009;

Keizer, 2015), which translates into research methodologies (Eriksson, Kovalainen, 2016), and then its interpretation.

The division into quantitative and qualitative methods is the basic division of methods in research sciences (Kam, Lai, 2018). Quantitative methods make it possible to process with statistical tools (Ferreira, Cova, Spencer, Proença, 2017), but the results are more general. Qualitative methods allow for in-depth observation of special cases (Gerring, 2007), but the data cannot be generalised per population. In order to select the correct methods and techniques for the study, it is necessary to start with the objectives of the study and research questions (Jap, Anderson, 2007). Correctly formulated goals and questions (Wade, 2004; Williamson, 2009) impose a certain minimum set of research methods, without which it will not be possible to answer the questions asked.

Surveys are the basic technique used in quantitative research (Miterev, Mancini, Turner, 2017). Depending on the methods of reaching the respondent, the following types of quantitative research can be distinguished (Lee, Johnsen, 2012): CATI (Computer Assisted Telephone Interview), PAPI (Paper and Pencil Interview), CAPI (Computer Assisted Personal Interview), CAWI (Computer Assisted Web Interview). Obviously, it is desirable and necessary to expand this set in order to multiply the methods (i.e. the so-called triangulation) and increase the quality of evaluation. Triangulation consists in the multiplication of methods, techniques and data sources) in order to confront the received information and summarise it, and allows to reduce measurement errors (Maylor, Turkulainen, 2019) and increase the quality of the study. The test results then become less error-prone and therefore more reliable. In order to increase the resistance of the research to errors, we can multiply: research methods and techniques, information sources, types of data and analytical techniques, explanatory theories, the number of people carrying out the research (Restuccia, Legoux, 2019). This makes it possible to detect and correct examination errors by crossing different viewpoints.

Testing on a sample can never absolve the researcher from error. The choice of the error size depends on the subject of the study and the level of accuracy at which we want to study a given population (Samimi, Sydow, 2021). To minimise the error it is necessary to increase the sample size. Thus, the choice of the tolerable error rate is usually a trade-off between the cost and quality of the audit. In management sciences, if there is a mistake or a gap in the study, the questionnaire/interview is usually eliminated as an incomplete data set.

Currently, in the field of social sciences a mixed approach (Molina-Azorin, 2016) is recommended, including management and quality sciences (Harrison, Reilly, Creswell, 2020), among others in the area of strategic management (Molina-Azorin, 2011; 2012). Research "involving at least one quantitative method (designed for collecting numbers) and one qualitative method (designed for collecting words) in which no type or method of research is inherently tied to any particular research paradigm" (Greene, Caracelli, Graham, 1989, p. 256) are very important. Therefore, attention was paid to the elimination of future errors and gaps and the use of a mixed approach, which we should conduct when researching study areas in

which the results so far have been ambiguous and/or fragmentary, and therefore there is no reliable and holistic knowledge (Johnson, Onwuegbuzie, 2004), or else identification of complex research problems and phenomena takes place (Molina-Azorin, 2012; 2016; Bazeley, 2015; Hong, Pluye, 2019). Therefore, in order to avoid the error of the research sample, it is necessary to use a mixed approach, because the importance of research, its results and conclusions drawn on their basis increases (Molina-Azorin, 2011; Gibson, 2017; Harrison, Reilly, Creswell, 2020), as a result of which generalisation becomes more authorised (Denscombe, 2008), the interpretations are more accurate (Gibson, 2017) and a more in-depth and more comprehensive interpretation of the results takes place, which allows for a more comprehensive picture of the phenomenon under study (Molina-Azorin, 2011).

Surveys have played a huge role, but when used in a mixed approach, they affect the originality of the course of research and the results obtained as a result, as well as the generally achieved level of methodological rigour (Molina-Azorin, 2012; Bazeley, 2015). Moreover, the applied triangulation of data and methods (Chen, 2006), but also the combination of multidimensional advantages of the quantitative and qualitative approach (Chen, 2006; Molina-Azorin, 2012; 2016; Harrison, Reilly, Creswell, 2020) influence the interpretation and elimination of research gaps.

The problem of filling gaps in data is a widely considered issue. Due to its numerous applications, it is of interest to many researchers and new methods of filling in missing data are constantly being developed (Zhang, 2016; Jerez, 2010; Bertsimas, 2017). Among the many approaches, the methods using fuzzy sets (Kiersztyn, 2020) or machine learning (Whitehead, 2019) deserve special attention. In many cases, missing data are a special case of anomalies or outliers that can also be detected and classified (Kiersztyn, 2020; 2020) and then modified to be more consistent with the rest of the dataset.

This paper proposes an innovative combination of many known machine learning techniques. The starting point was the unusual use of association techniques to define the explanatory variables. Based on the relationship between the individual variables resulting from the basket analysis, the process of supplementing data gaps was carried out with the use of selected techniques used for classification. The paper considers a special type of data, which are the results of surveys typical for research in the field of economics and management. The proposed solution, due to its intuitiveness and ease of adaptation, can be successfully used by researchers who do not have much experience in the use of machine learning.

The structure of the work is as follows. Section II describes the dataset on which the research was performed. Then, Section III provides a description of the methodology used for filling data gaps. In the next Section IV, the results of numerical experiments are presented. The last Section V is devoted to conclusions and future work directions

2. Description of the data set analysed

The research used in this article is part of a research project under the name of "Teal organisations in Economy 4.0" (Rzepka, 2020; 2021; 2022). The project involves conducting research in Poland (Maciaszczyk et al., 2023; Miśkiewicz et al., 2021), and in selected countries of the world (USA, Georgia, Slovakia, Brazil, England, Romania, Czech Republic, Ukraine, Spain). The research is conducted in stages and includes a pilot study (N = 300), core research (N = 300 PL and 330 different countries) and repeat research (N = 300). It was a relatively extensive survey distributed among top management from different countries. The questionnaire consisted of parts devoted to different topics, like general information, innovation and technology, relationship, social capital, knowledge and information, trust, structure, organisational culture, associations and personal profile), and each part had 5 to 7 questions. The aim of the survey was to examine the extent to which contemporary companies cooperate with consumers and exhibit the characteristic features of Teal organisations (Rzepka, 2023; Miśkiewicz et al., 2021) as well as the extent to which these features influence various aspects of the operation of the company, including its ability to innovate. It should be noted that agility remains an inherent characteristic of Teal organisations (Rzepka, 2023).

The study was divided into stages that includes a pilot study (May-June 2020), stage I: quantitative study (July-September 2020) and stage II: December 2020 – January 2021. One top management representative of each enterprise was asked to participate in the survey. The choice of enterprises resulted from the SMEs' availability. Moreover as a part of the survey 15 structured direct interviews were conducted in Polish companies. These interviews constitute the pilot studies for further stages of direct interviews in other countries participating in the research.

It is worth presenting the results of research from the aforementioned project. The study has been and will be conducted according to the principles and standards developed by the Network on Development Evaluation of the OECD Development Assistance Committee (DAC). The following work was carried out in the course of the study – Desk research; IDI (Individual In-Depth Interview), and questionnaire study with selected groups of people using Computer-Assisted Web Interview (CAWI) and Paper and Pencil Interview (PAPI) techniques.

At each stage, a sample of 300 respondents from various micro, small and medium-sized enterprises, with different levels of coverage and size, was selected (Table 1).

Table 1.
Scope of respondents

		Pilot	I stage	II stage
Predominant mode of the company's operation (%)	Commerce	19.7	10.3	16.6
	Production	17.6	13.9	28.0
	Services	62.7	75.8	55.4
Scope of the company's activity (%)	Local	19.3	22.1	9.4
	Regional	8.2	12.4	27.3
	National	30.0	23.0	22.0
	International	42.5	42.4	41.3
Number of employees (%)	0-9	13.7	9.1	3.3
	10-49	24.0	18.8	55.1
	50-249	15.5	28.2	16.6
	250-999	18.5	20.9	16.6
	1000 – and more	28.3	23.0	8.4

The survey included top-level employees from companies with 50-249 employees (28.3%) with an international scope (42.5%), while the largest number of respondents in Stage III represented service companies (55.4%), located in multinationals (41.3%) and employing an average of 10-49 people (55.1%). Within the industry, as many as 11.4% represented the IT industry during the period under review.

3. Methodology for filling gaps in data

When supplementing the missing survey data, the characteristics of the analysed set should be taken into account. The most intuitive way to fill in the missing data in the questionnaires is to use the mean or median of the answers determined on the basis of the available data. This approach does not take into account the nature of the question, but only estimates the missing value based on the basic properties of the distribution of answers. This approach, despite its simplicity, can yield relatively good results. In the proposed approach, we will use a slightly more complex approach that produces much better results. The starting point of the proposed solution is the application of association rules to fill in missing data. In the case of survey data based on the Likert scale, there is a finite number of combinations of possible answers. Therefore, it is possible to determine basic measures describing the dependencies occurring in association rules for all possible sets of answers. The most important measure in the case under consideration seems to be the trust set by the formula

$$confidence(X,Y) = \frac{supp(X \cup Y)}{supp(X)} \tag{1}$$

where the support is set by the formula

$$supp(X) = \frac{\#X}{N} \tag{2}$$

which denotes the percentage of rules where the value of X occurs.

In order to increase the clarity of considerations, let us introduce the following symbols and the necessary assumptions. Assume that the dataset analysed consists of a set of N questions $Q_1, Q_2, \dots, Q_N$, each of which uses a K graded Likert scale. It is possible to make the number of available answers dependent on the question number and thus allow the use of a different scale for individual questions. However, such an option will significantly reduce the transparency of the record, and on the other hand will not bring additional benefits. The proposed innovative method does not depend on the number of questions, while the introduction of additional indices will reduce the transparency of the theoretical part. To summarise, for the n-th of N questions, possible answers belong to the set $\{X_1^n, X_2^n, \dots, X_K^n\}$. Most often, the values $X_j^n = j, j = 1,2, \dots, K$, however, it is possible to introduce a different scale and we will not introduce this limitation.

On the basis of the available data, it is possible to determine the confidence value (1) for each combination of questions and possible answers, i.e.

$$confidence(X_j^n, X_i^m)$$

for $i, j \in \{1,2, \dots, K\}, n, m \in \{1,2, \dots, N\}$.

The values obtained in this way are the starting point for filling gaps in the data. If for a certain respondent there is no answer in the question Q_n, then the remaining available data for that respondent should be analysed and on their basis the chances of the occurrence of each of the possible answers $\{X_1^n, X_2^n, \dots, X_K^n\}$ should be estimated. More precisely, based on the answers to the remaining questions, the values of the confidence measures for the respective combinations are summed up. As a result, the CON vector of aggregated values of the confidence measure is obtained

$$CON = [con(X_1^n), con(X_2^n), \dots, con(X_K^n)] \tag{3}$$

where

$$con(X_i^n) = \sum_{j=1}^{K} \sum_{m \neq n}^{N} confidence\left(X_j^n, X_i^m\right) \tag{4}$$

For the suggested value, choose the answer for which the value of the aggregate confidence measure is the highest. The results obtained with this novel application of the classic machine learning tool are very promising, as shown in the experimental section.

4. Experimental results

In order to test the effectiveness of the proposed method, a series of numerical experiments were performed on the data set discussed in Chapter 3. More precisely, 40 questions were selected for the analysis, for which there is an economic justification for the existence of relationships between the individual answers. Each of the analysed questions made it possible to select one of the 5 answers according to the Likert scale as the answer. In the first step,

the correlation between the individual variables was determined for the available data (cf. Fig. 1). The values of the linear correlation coefficient will be used in the following to weight confidence measures between individual responses.

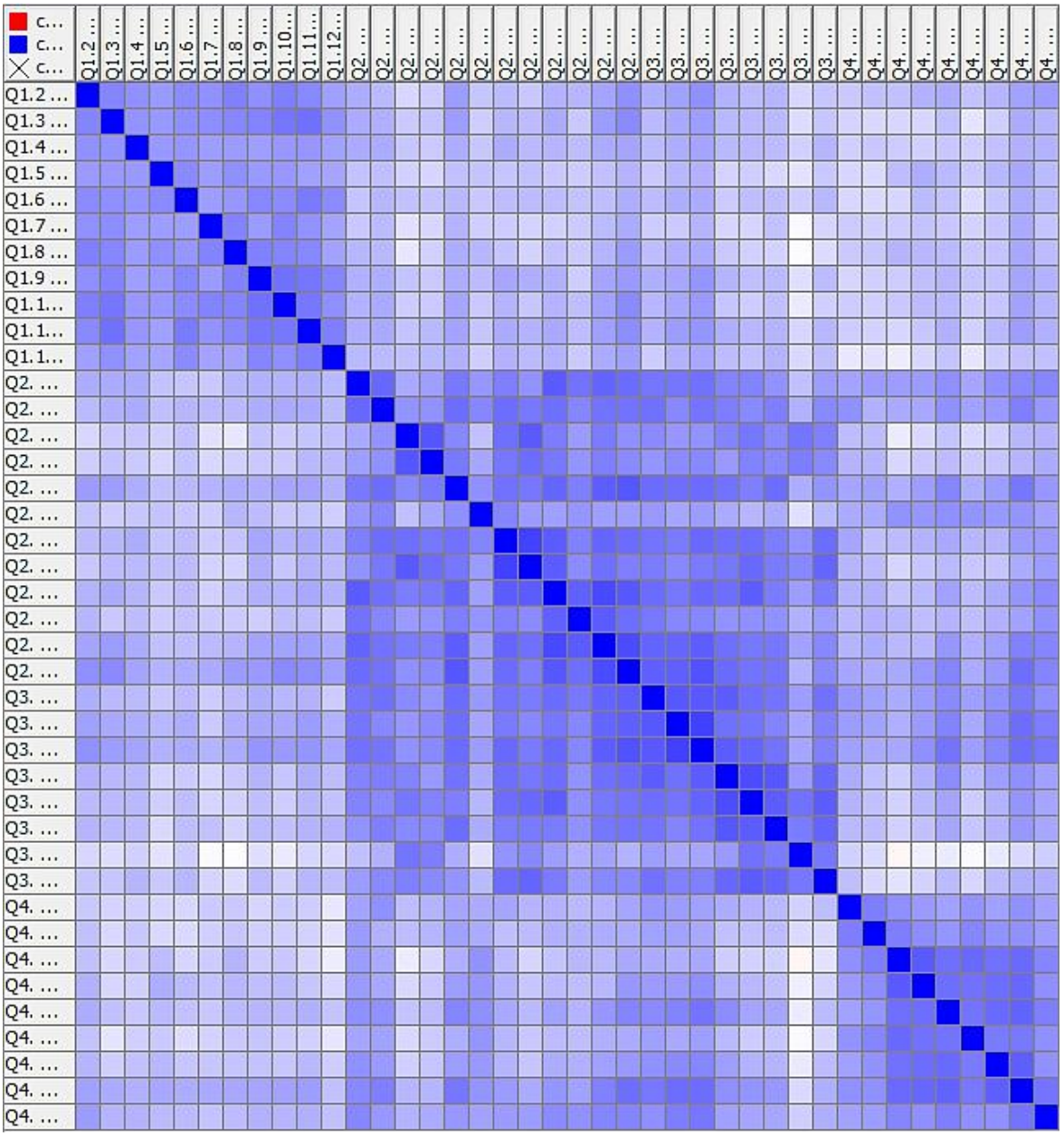

Figure 1. The level of Pearson's linear correlation between the considered variables corresponding to the individual questions in the survey.

Analysing the results presented in Fig. 1, it can be seen that for some pairs of questions there is a relatively high relationship, which is well illustrated by a linear correlation. This observation will be used later in the experimental section. Apart from the correlation between the individual ones, before the artificial introduction of gaps, the values of the confidence measure were determined for all possible answers and questions. Thus, a square CONF matrix with dimensions of 200x200 was determined. It should be noted at this point that this matrix is obviously not symmetrical. Having the necessary values, it was possible to test

the effectiveness of the proposed solution by introducing random gaps in the data corresponding to the failure of the respondent to answer a given question.

Data gaps were generated according to a uniform distribution in a two-step manner. In the first step, the respondent (a row in the data table) was selected, and then the question (a column in the data table). This draw was repeated each time a predetermined number of repetitions, while the possibility of repetitions was allowed, which was necessary in the case of indicating a large number of repetitions. The empirical distribution of the randomly selected positions of the gaps is shown in Fig. 2.

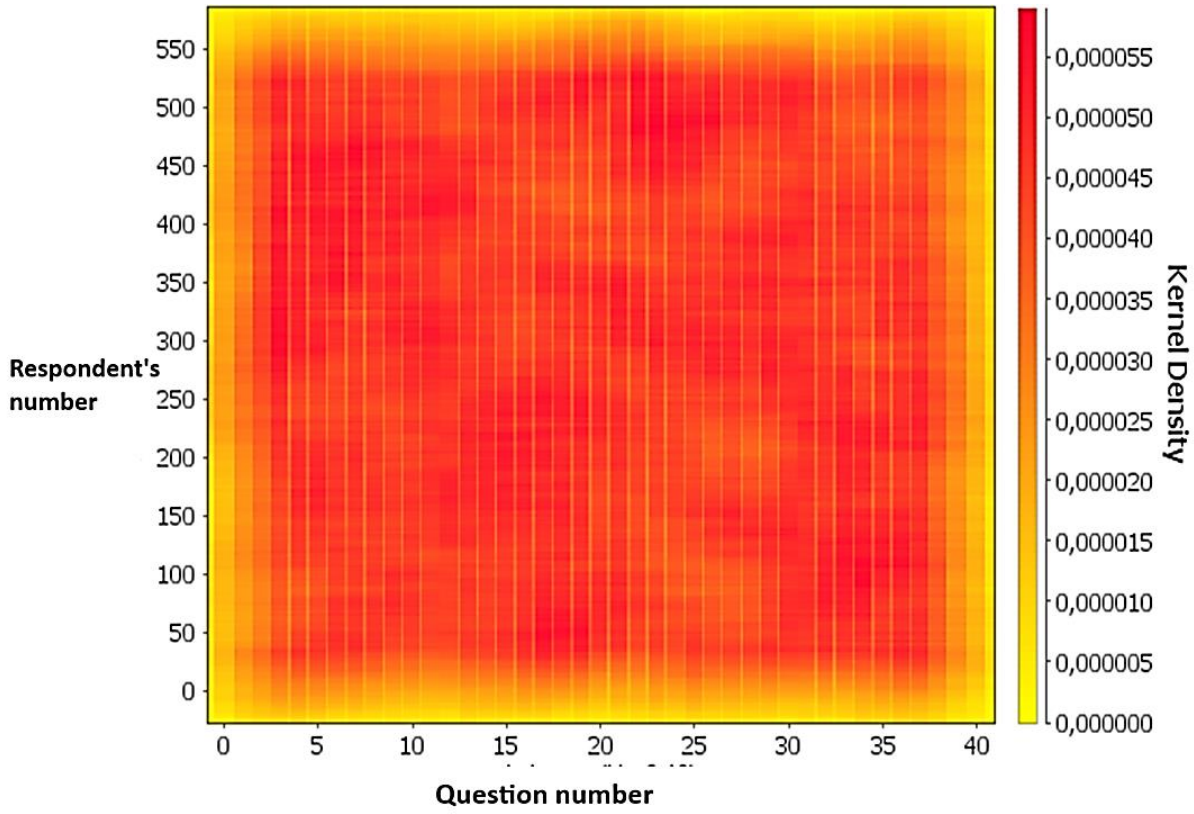

Figure 2. Distribution of items of randomly selected missing data in the analysed data set.

In the next step, for each of the gaps, the vector of aggregated confidence measures was determined for the remaining 39 questions in two ways. In the first approach, the sum of the values of confidence measures was calculated after the appropriate pairs of responses, i.e. in other words, the suitable values from the CONF matrix were summed. Then the obtained values were normalised in such a way that the sum of the components of the vector CON was 1. The position of the highest value of the vector CON corresponds to the most probable value that has been removed. In the second approach, a slight modification was introduced, consisting in the use of additional weight when summing the appropriate elements of the CONF matrix. The measure used was the value of the correlation coefficient between the individual questions. Two competing solutions were thus obtained.

Very satisfactory results were obtained in the case of generating 1000 gaps and applying the simplified version of missing data filling using only association measures. In 643 cases, correct restoration of deleted values was obtained. The distribution of differences between the real value and the result obtained with the proposed method is shown in Fig. 3.

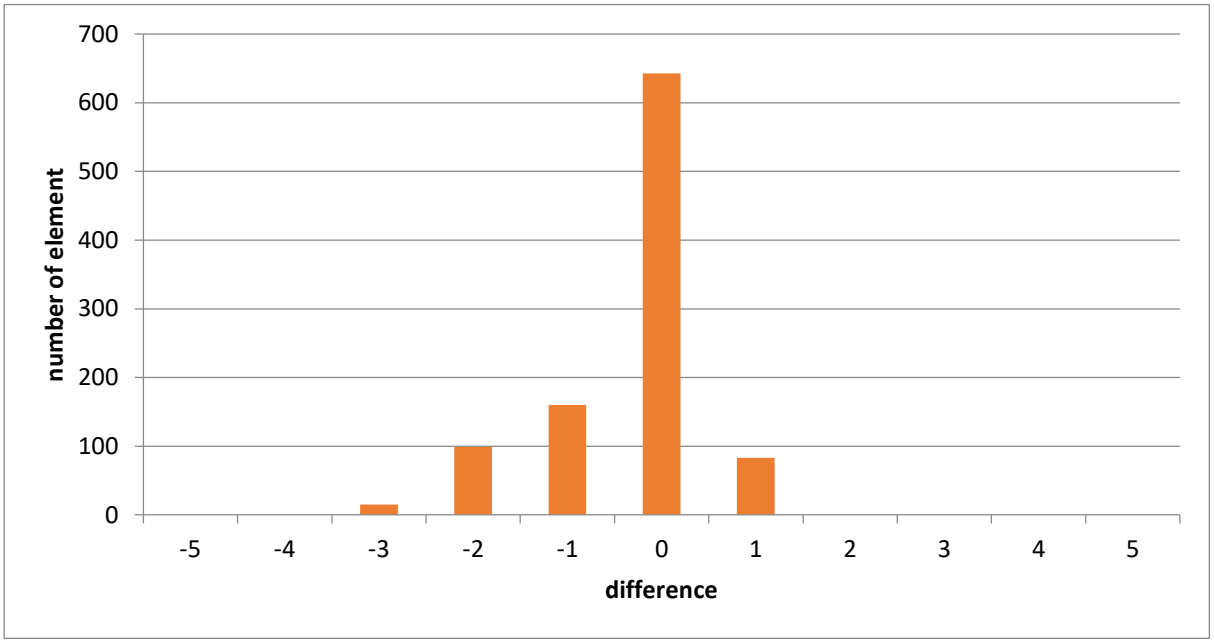

Figure 1. Distribution of differences.

The results presented in Fig. 3 show that when the correct answer was not given, the difference between the real value and the model was small and only in 15 was it greater than 2.

If the proposed method of supplementing missing data is enriched by applying an additional weight corresponding to the correlations between the questions, even better results are obtained. In 706 cases, the deleted value was perfectly reproduced and the differences between the empirical and theoretical values were even smaller, cf. Fig. 4.

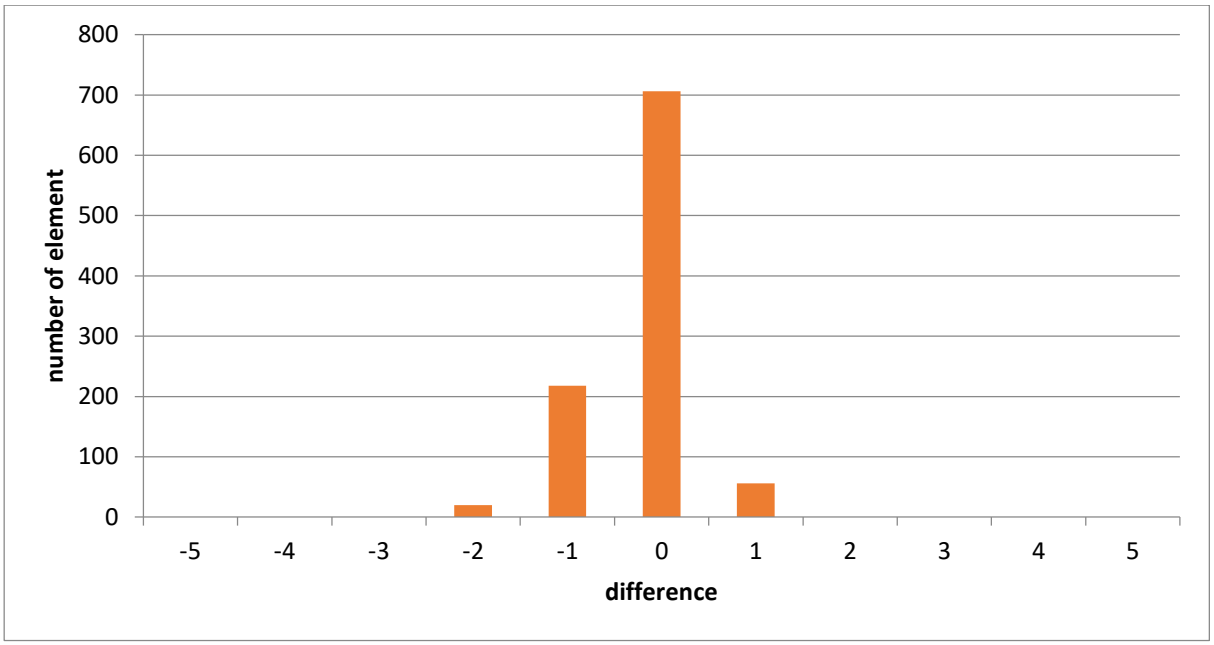

Figure 2. Density of differences.

Additionally, it is worth noting that if we consider the second of the suggested answers, then in the typed pair of values with a probability of 0.965 there is actually the deleted value. This observation suggests the use of additional analyses involving the addition of machine learning techniques to determine the removed value based on the values of the CON vector.

It turns out that the use of additional machine learning tools can significantly increase the effectiveness of the proposed solution. For example, for the data containing the CON vector values for individual 1000 artificially generated missing data, several well-known classification methods were used (Kiersztyn, 2021a). The methods implemented in KNIME were selected for the analysis, namely FuzzyRule (FR) (Berthold 2003), RandomForest (RF) (Pal, 2005), TreeEnsamble (TE) (Coppersmith, 1999), Gradient BoostedTree (GBT) (Friedman, 2002). As is generally known, the random allocation of elements to the training set has a huge impact on the effectiveness of the classification (Kiersztyn, 2021b), in order to eliminate the impact of the randomness of the division within each division of the available data on the training and test sets, 10 independent repetitions of the test were performed. The averaged results for different partitions are presented in Table 2.

Table 2.
Classification efficiency depending on the size of the training set

Percentage of elements in the training set	FR	RF	TE	GBT	DT
5	83.624	97.263	97.368	94.421	88.526
10	99.204	99.111	99.111	99	93.333
15	97.934	98.706	98.353	96.706	97.765
20	99.493	99.125	99.625	96.750	96.500
25	99.454	99.867	99.867	99.867	99.867
30	100	100	100	99.714	99.143
35	100	100	100	100	100

For a greater percentage of elements in the training set, all values are equal to 100 for each method. Analysing the results presented in Table 1, we can see that, regardless of the classification method used, we obtain very high classification efficiency for small teaching sets. It turns out that in many cases the algorithm was able to correctly classify the remaining values on the basis of 30% of the elements classified to the training set. It should be noted here that even in the event of an incorrect classification, the difference between the actual state and the theoretical value is small, as evidenced by the example of the RF classification result for 5% of the elements in the training set is presented in Table 2.

Table 3.
An example of the results of classification by the RF Method

Real value/Predicted value	1	2	3	4	5
1	77	16	0	0	0
2	1	142	16	0	0
3	0	0	222	18	0
4	0	0	0	274	11
5	0	0	0	8	165

It can be seen that the prediction differs at most by one value from the actual state and usually the RF method, as well as the others, overestimate the predicted value. It should be noted here that the above values presented in Table1 and Table2 are obtained when the analyses are performed based on the value of the CON vector and the number of the column in which the value was removed. In other words, during the analysis, the information was available from which question the value was removed. Information about the respondent's number has not been added to the set of explanatory variables. If the information about the number of the question from which the analysed value has been removed is omitted, the effectiveness changes slightly. By repeating the entire classification process for a smaller number of explanatory variables, the results presented in Table 3 were obtained.

Table 4.
Efficiency of individual methods on a limited number of explanatory variables

Percentage of elements in the training set	FR	RF	TE	GBT	DT
5	94.842	95.895	94.842	89.895	89.368
10	99.887	95.556	96.222	97.333	95
15	99.647	98.588	99.059	98.588	96.706
20	99.757	99.500	99.500	99.500	97.625
25	99.867	99.867	99.867	99.867	97.467
30	100	100	100	100	100

Comparing the results presented in Table 1 and Table 3, it should be noted that in the case of the FuzzyRule (FR) method, reducing the number of variables, i.e. limiting information, surprisingly increased efficiency. Similarly, in the case of DecisionTree (DT), removing the number of the analysed question information tended to increase efficiency. The reasons for such

a state can be found in the randomness of the data division into the training and test sets. Nevertheless, due to the repetition of experiments for the two compared approaches, the impact of randomness seems to be limited. The differences between the effectiveness of individual methods with and without available information about the analysed question (column) are shown in Fig. 5.

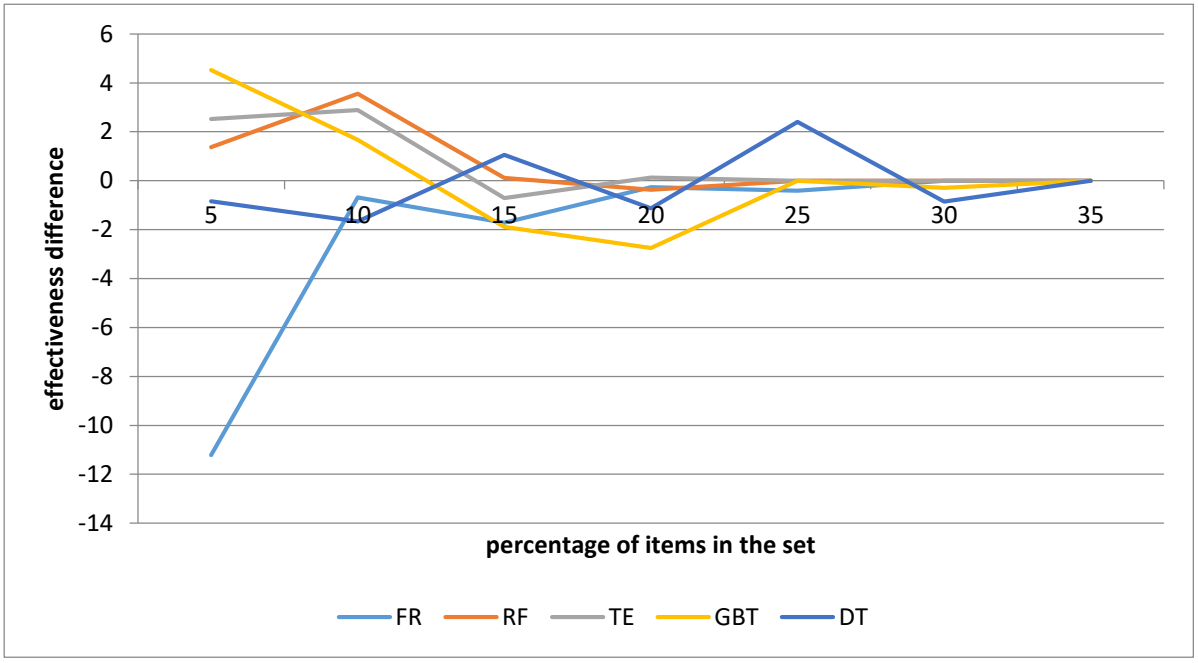

Figure 5. Differences in the effectiveness of individual methods for different explanatory variables.

Analyzing the results presented in Fig. 5, we see that as the number of elements in the training set increases, the efficiency increases. This is an obvious phenomenon, but achieving full compliance with 30% of the elements in the training set is a very good result. Moreover, there are slight differences between the compared methods.

5. Conclusion and future work

The proposed innovative approach to filling gaps in data in questionnaire surveys has a number of practical applications. In many cases, the results of the surveys carried out contain gaps in the data due to various factors. Very good results have been achieved thanks to the appropriate use of machine learning tools, in particular thanks to the skillful combination of different approaches. In the experimental section, the effectiveness of the proposed solution was confirmed, showing that on the basis of a small number of training sets, it is possible to obtain 100% correctness in recreating artificially generated data gaps. Moreover, the obtained results indicate a high potential for conducting interdisciplinary research and supporting researchers in the field of management with the great possibilities of artificial intelligence tools.

It is planned to extend the research to survey questions in which the answers are not limited to the Likert scale. In addition, work is underway on the use of other classification techniques and testing of the obtained solutions on other datasets that do not necessarily describe survey research.

References

1. Bazeley, P. (2015). Writing up multimethod and mixed methods research for diverse audiences. In: S.S. Hesse-Biber, R.B. Johnson (Eds.), *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry* (pp. 296-313). Oxford University Press.
2. Bennett, A., Elman, C. (2006). Qualitative research: Recent developments in case study methods. *Annu. Rev. Polit. Sci., vol. 9,* pp. 455476.
3. Berthold, M.R. (2003). Mixed fuzzy rule formation. *Int. J. Approx. Reason., vol. 32, no. 2-3,* pp. 67-84.
4. Bertsimas, D., Pawlowski, C., Zhuo, Y.D. (2017). From predictive methods to missing data imputation: an optimization approach. *J. Mach. Learn. Res., vol. 18, no. 1,* pp. 7133-7171.
5. Boumans, M., Davis, J.B. (2015). *Economic methodology: Understanding economics as a science*. Macmillan International Higher Education.
6. Burke, C.M., Morley, M.J. (2016). On temporary organizations: A review, synthesis and research agenda. *Hum. Relat., vol. 69, no. 6,* pp. 1235-1258.
7. Coppersmith, D., Hong, S.J., Hosking, J.R. (1999). Partitioning nominal attributes in decision trees. *Data Min. Knowl. Discov., vol. 3, no. 2,* pp.197-217.
8. Denscombe, M. (2008). Communities of practice: A research paradigm for the mixed methods approach. *J. Mix. Methods Res., vol. 2, no. 3*, pp. 270-283.
9. Düppe, T. (2011). How economic methodology became a separate science. *J. Econ. Methodol., vol. 18, no. 2,* pp. 163-176.
10. Eriksson, P., Kovalainen, A. (2015). *Qualitative methods in business research: A practical guide to social research.* Sage.
11. Farmer, J.D. (2013). Hypotheses non fingo: Problems with the scientific method in economics. *J. Econ. Methodol., vol. 20, no. 4,* pp. 377-385.
12. Ferreira, F.N.H., Cova, B., Spencer, R., Proença, J.F. (2017). A phase model for solution relationship development: a case study in the aerospace industry. *J. Bus. Ind. Mark.*
13. Friedman, J.H. (2002). Stochastic gradient boosting. *Comput. Stat. Data Anal., vol. 38, no. 4,* pp. 367-378.
14. Gerring, J. (2006). *Case study research: Principles and practices*. Cambridge University Press.
15. Gibson, C.B. (2017). Elaboration, generalization, triangulation, and interpretation:

On enhancing the value of mixed method research. *Organ. Res. Methods, vol. 20, no. 2,* pp. 193-223.

16. Greene, J.C., Caracelli, V.J., Graham, W.F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educ. Eval. Policy Anal., vol. 11, no. 3*, pp. 255-274.
17. Hands, D.W. (2004). Pragmatism, knowledge, and economic science: Deweyan pragmatic philosophy and contemporary economic methodology. In: E.L. Khalil (Ed.), *Pragmatism, Knowledge, and Economic Science* (pp. 255-270). London/New York: Routledge.
18. Harrison, R.L., Reilly, T.M., Creswell, J.W. (2020). Methodological rigor in mixed methods: An application in management studies. *J. Mix. Methods Res., vol. 14, no. 4,* pp. 473-495.
19. Hong, Q.N., Pluye, P. (2019). A conceptual framework for critical appraisal in systematic mixed studies reviews. *J. Mix. Methods Res., vol. 13, no. 4*, pp. 446-460.
20. Jap, S.D., Anderson, E. (2007). Testing a life-cycle theory of cooperative interorganizational relationships: Movement across stages and performance. *Manage Sci., vol. 53, no. 2*, pp. 260-275.
21. Jerez, J.M., Molina, I., García-Laencina, P.J., Alba, E., Ribelles, N., Martín, M., Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artif. Intell. Med., vol. 50, no. 2*, pp. 105-115.
22. Johnson, R.B., Onwuegbuzie, A.J. (2004). Mixed methods research: A research paradigm whose time has come. *Educ. Res., vol. 33, no. 7*, pp. 14-26.
23. Kam, B.H., Lai, M.K. (2018). Buyer-supplier exchange relationship: How do exchange partners behave across the relationship life-cycle? *Transp. Res. E: Logist. Transp. Rev., vol. 113*, pp. 239-257.
24. Karczmarek, P., Kiersztyn, A., Pedrycz, W., Al, E. (2020). K-means-based isolation forest. *Knowl.-Based Syst., vol. 195*, p. 105659.
25. Keizer, P. (2015). *Multidisciplinary Economics: A Methodological Account*. Oxford: University Press.
26. Kiersztyn, A., Karczmarek, P., Kiersztyn, K., Pedrycz, W. (2020). *The concept of detecting and classifying anomalies in large data sets on a basis of information granules*. 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1-7.
27. Kiersztyn, A., Karczmarek, P., Kiersztyn, K., Pedrycz, W. (2021). *Detection and classification of anomalies in large data sets on the basis of information granules*. IEEE Trans. Fuzzy Syst.
28. Kiersztyn, A., Karczmarek, P., Łopucki, R., Pedrycz, W., Al, E., Kitowski, I., Zbyryt, A. (2020). *Data imputation in related time series using fuzzy set-based techniques*. 2020 IEEE international conference on fuzzy systems (FUZZ-IEEE). IEEE, pp. 1-8.
29. Kiersztyn, A., Kiersztyn, K., Karczmarek, P., Kaminski, M., Kitowski, I., Zbyryt, A., Łopucki, R., Pitucha, G., Pedrycz, W. (2021). *Classification of complex ecological objects with the use of information granules.* 2021 IEEE International Conference on Fuzzy

Systems (FUZZ-IEEE). IEEE, pp. 1-6.
30. Kiersztyn, A., Łopucki, R., Kiersztyn, K., Karczmarek, P., Powroźnik, P., Czerwiński, D., Pedrycz, W. (2021). *A comprehensive analysis of the impact of selecting the training set elements on the correctness of classification for highly variable ecological data*. 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE, pp. 1-6.
31. Kiersztyn, K. (2021). Intuitively adaptable outlier detector. *Stat. Anal. Data Min.: ASA Data Sci. J.*, pp. 1-17.
32. Lee, C.-J., Johnsen, R.E. (2012). Asymmetric customer–supplier relationship development in Taiwanese electronics firms. *Ind. Mark. Manag., vol. 41, no. 4*, pp. 692-705.
33. Maciaszczyk, M., Makieła, Z., Miśkiewicz, R. (2023). Industry 5.0: A New Reality, New Challenges. In: A. Rzepka (Ed.), *Innovation in the Digital Economy New Approaches to Management for Industry 5.0* (pp. 51-61). London/New York: Routledge. ISBN 978-1-03-246993-5.
34. Mäki, U. (2009). *The methodology of positive economics: Reflections on the Milton Friedman legacy.* Cambridge University Press.
35. Maylor, H., Turkulainen, V. (2019). The concept of organisational projectification: past, present and beyond? *Int. J. Manag. Proj. Bus*.
36. Miśkiewicz, R. (2020). Efficiency of Electricity Production Technology from Post-Process Gas Heat: Ecological, Economic and Social Benefits. *Energies, 13, 6106*.
37. Miśkiewicz, R., Rzepka, A., Borowiecki, R., Olesiński, Z. (2021). Energy efficiency in the industry 4.0 era: Attributes of teal organisations. *Energies, vol. 14, no. 20*, p. 6776.
38. Miśkiewicz, R., Rzepka, A., Borowiecki, R., Olesiński, Z. (2021). Energy Efficiency in the Industry 4.0 Era: Attributes of Teal Organisations. *Energies, 14, 6776*.
39. Miterev, M., Mancini, M., Turner, R. (2017). Towards a design for the project-based organization. *Int. J. Proj. Manag., vol. 35, no. 3*, pp. 479-491.
40. Molina-Azorín, J.F. (2011). The use and added value of mixed methods in management research. *J. Mix. Methods Res., vol. 5, no. 1*, pp. 7-24.
41. Molina-Azorín, J.F. (2012). Mixed methods research in strategic management: Impact and applications. *Organ. Res. Methods, vol. 15, no. 1*, pp. 33-56.
42. Molina-Azorín, J.F., López-Gamero, M.D. (2016). Mixed methods studies in environmental management research: Prevalence, purposes and designs. *Bus. Strategy Environ., vol. 25, no. 2*, pp. 134-148.
43. Molina-Azorín, J.F., López-Gamero, M.D., Pereira-Moliner, J., Pertusa-Ortega, E.M. (2012). Mixed methods studies in entrepreneurship research: Applications and contributions. *Entrepreneurship Reg. Dev., vol. 24, no. 5-6*, pp. 425-456.
44. Pal, M. (2005). Random forest classifier for remote sensing classification. *Int. J. Remote Sens., vol. 26, no. 1*, pp. 217-222.

45. Restuccia, R., Legoux, R. (2019). B2b relationships on the fast track: An empirical investigation into the outcomes of solution provision. *Ind. Mark. Manag., vol. 76,* pp. 203-213.
46. Rzepka, A. (2020). Turkusowe organizacje w Gruzji. In: Z. Olesiński (Ed.), *Składniki turkusowych organizacji* (pp. 293-306). Warszawa: Difin.
47. Rzepka, A. (2021). Teal organizations in times of Industry 4.0. *Eur. Res. Stud. J., vol. 24,* pp. 60-71.
48. Rzepka, A. (2022). *Self-management, Entrepreneurial Culture, and Economy 4.0: a Contemporary Approach to Organizational Theory Development*. London/New York: Routledge.
49. Rzepka, A. (2023). *Innovation in the Digital Economy New Approaches to Management for Industry 5.0.* London/New York: Routledge.
50. Samimi, E., Sydow, J. (2021). Human resource management in project based organizations: revisiting the permanency assumption. *Int. J. Hum. Resour. Manag., vol. 32, no. 1*, pp. 49-83.
51. Whitehead, T.M., Irwin, B.W., Hunt, P., Segall, M.D., Conduit, G.J. (2019). Imputation of assay bioactivity data using deep learning. *J. Chem. Inf. Model., vol. 59, no. 3*, pp. 1197-1204.
52. Williamson, O.E. (2009). Pragmatic methodology: a sketch, with applications to transaction cost economics. *J. Econ. Methodol., vol. 16, no. 2,* pp. 145-157.
53. Zhang, Z. (2016). Missing data imputation: focusing on single imputation. *Ann. Transl. Med., vol. 4, no. 1.*