symmetry

MDPI

Article

Information Reliability Assessment and Filtering for Decision Making. Case Study of Lithuanian and China Students' Life Goals Survey Data

Aleksandras Krylovas [1], Natalja Kosareva [1] and Stanislav Dadelo [2,*]

[1] Department of Mathematical Modelling, Vilnius Gediminas Technical University, Sauletekio al. 11, LT-10223 Vilnius, Lithuania; aleksandras.krylovas@vilniustech.lt (A.K.); natalja.kosareva@vilniustech.lt (N.K.)
[2] Department of Entertainment Industries, Vilnius Gediminas Technical University, Sauletekio al. 11, LT-10223 Vilnius, Lithuania
* Correspondence: stanislav.dadelo@vilniustech.lt; Tel.: +370-69882824

Abstract: This article presents the methodology and tools to evaluate the reliability of quantitative sociological research data. The problem of filtering unreliable data is usually solved by statistical methods. This article proposes an improved method for filtering unreliable data. In this case, the statistical methods are not applied to the initial data but the value of the distance function between the two preferences. This allows for the disclosure of conflicting or erroneous data. Calculation of the distance between two preferences and prioritisation of life goals are based on binary relation theory, where the properties of symmetry (antisymmetry) are very important. The article presents a case study on 11 life goals evaluation and ranking by Lithuanian and China students. The study revealed that the China student data filtered at least twice as much as the Lithuanian student data, i.e., they are less reliable. The filtered data show that students of both countries ranked the most and the least important life goals in a very similar way with minimum deviations detected in the ranking results.

Keywords: information filter; preference incompatibility assessment; data reliability; alternatives ranking; life goals; decision making

Citation: Krylovas, A.; Kosareva, N.; Dadelo, S. Information Reliability Assessment and Filtering for Decision Making. Case Study of Lithuanian and China Students' Life Goals Survey Data. *Symmetry* **2021**, *13*, 2452. https://doi.org/10.3390/sym13122452

Academic Editor: Jian-Qiang Wang

Received: 21 November 2021
Accepted: 17 December 2021
Published: 20 December 2021

1. Introduction

Information flows are increasing in all areas of human activity. All ongoing and planned activities are based on data [1]. Information provision is developed in three main directions: (1) provision of actual data for the management purposes; (2) data supply to automatic control systems; (3) provision of data to the users of various types of information (organizations, researchers, artists, writers, journalists, etc.) [2].

Lately, the focus has been placed on digitising information and processing [3]. One of the most critical tasks in processing and analysing large volumes of information is to meet the needs of the target audience [4,5]. Meeting the needs of the target audience is accompanied by an increase in the volume of information [6], and the increase in the volume of information leads to new challenges for information providers and consumers [7]. Collecting and sending information to the end user determines its quantitative and qualitative changes. When information is obtained from multiple sources, its quality, accuracy, and reliability undergo fundamental changes. Particular attention should be paid to individual-generated information that may have latent hazards, such as gaps in the information provided (deleted and missing information, truncated information), irrelevant copy–pasted or unrealistic information. This can lead to erroneous conclusions, interpretations, recommendations and form a false image of the subject under assessment [8].

The assessment of information reliability should ensure the degree of correspondence between existing knowledge about the object (within the framework of information models

adopted for problem-solving) and the real state of the object, i.e., it should ensure the adequacy of object or phenomenon representation of the real object or the actual state of affairs, and determine the degree of compliance [9].

The complexity of the information reliability assessment process is determined by the need to evaluate all information in such a way that the removal of insignificant or erroneous information (inaccuracies) does not obscure the true state of the object and does not distort the correct perception of a particular object or a phenomenon. Thus, the provision of information to the consumer rests on the development of objective and sustainable research process and should guarantee the identification and elimination of false or falsified information [10].

The ability to quickly evaluate the information content and assess its reliability, possibly filtering information noise, is a pervasive issue. Developing the algorithm to carry out this process successfully and improving its tool makes it possible to minimise the inevitable error. This is particularly important in decision making [11]. Decision-makers should personally evaluate the information associated with the problem being addressed. It is necessary to develop the right tools for this task. If the decision-maker does not personally evaluate the information related to the problem being solved, the accuracy of the assessment and, consequently, the effectiveness of solving the problem may be significantly reduced. The process of information collection and selection becomes critical, and it is necessary to take all possible measures to identify misleading information at this stage [12]. Failure to follow this principle can lead to erroneous decisions with unintended consequences.

The range of information confidence levels is extensive, from absolute reliability to absolute uncertainty. This statement can be verified by assessing consumer behaviour [13]. On the other hand, reliability assessment can be influenced by subjective factors (individual and group) that significantly affect information reliability [14]. In the process of information collection and evaluation, it is important to formulate the description of the evaluated object properly, reliably so that it is clearly understandable to the assessors [15]. Thus, assessing the reliability of information is a multiobjective task.

Galvin and Seawright [16] proposed the selection of extreme cases for the treatment variable (the so-called "surprising causes"), net of the statistical influence of the set of known control variables. Schanze [17] analysed the data quality of older-age and cognitively impaired respondents. Researchers tend to label residents in retirement and nursing homes as hard-to-interview and exclude them from most social surveys.

The questionnaire is one of the most widely used data collection tools. Surveys aim to obtain reliable information, and the reliability of survey data is a significant factor in the research methodology [18,19]. The reliability of the survey data is usually equated with the validity of the questionnaire tool, but not enough attention is paid to assessing the reliability of the respondents' answers. This is an important interdisciplinary issue.

The article presents a case study on the life goals of students from Lithuania and China. The proposed method of filtering unreliable data was applied to four data sets. Finally, the priorities of life goals were determined by elaborating the decision-making theory. More information on the decision-making theory, as well as the examples of decision-making methods application for sociological and economic data, have been proposed in [20,21].

The aim of our research is to evaluate the reliability of quantitative sociological research data. To achieve the goal of the work, it was necessary to develop the tools and the criteria for assessing the reliability of the data. For this purpose, a preference comparison algorithm to identify poorly matched preferences was proposed. The algorithm is based on the binary relation theory and statistical methods, where the statistical characteristics of real preferences are compared with the characteristics of randomly generated preferences.

The rest of the article is organized as follows: Section 2 introduces preliminaries, defines notions of preference, the distance between two preferences, and shows how preferences can be compared. Section 3 describes 11 life goals (values) prioritised by the survey participants. The data filtering procedure based on statistical methods is proposed

in Section 4. The application of the data filtering procedure was demonstrated for the four data sets of Lithuanian and Chinese students. Section 4 also includes life goals prioritisation results. Section 5 provides the discussion and Section 6 conclusions.

2. Comparison of Preferences

It will be recalled [22] that the *relation* between the elements of set $A = \{a_1, a_2, \ldots, a_n\}$ is any subset $S = \{(a_i, a_j), i, j \in \{1, 2, \ldots, n\}\} \subset A \times A$. The relation S is called *transitive* when $(a_i, a_j)\&(a_j, a_k) \in S \Rightarrow (a_i, a_k) \in S$; *antisymmetric* when $(\forall i \neq j)\ (a_i, a_j) \in S \Rightarrow (a_j, a_i) \notin S$; and *complete* when $(\forall i \neq j)\ (a_i, a_j) \vee (a_j, a_i) \in S$. Antisymmetric, transitive, and complete (strict order) relation will be called *preference* and denoted by $\succ$. The relation S is called *reflexive* when $(\forall i)\ (a_i, a_i) \in S$; *symmetric* when $(a_i, a_j) \in S \Rightarrow (a_j, a_i) \in S$. Reflexive, symmetric, and transitive relation is called *equivalence* and denoted by $\sim$. Transitive and complete relation will be denoted by $\succcurlyeq$. Let us agree to write $a_i \sim a_j$ when $a_i \succcurlyeq a_j$ and $a_j \succcurlyeq a_i$. Notice that if $a_{i_1} \sim a_{i_2} \sim \ldots \sim a_{i_k}$, then $\sim$ is the equivalence relation in a set $A\prime = \{a_{i_1}, a_{i_2}, \ldots, a_{i_k}\}$. In this article, we will only consider such preferences, when there is at least one subset $U_{l_i''}$ of the set A that defines the equivalence relation $\sim$. Thus, there exists such a set A partition into blocks $T_{l_i'}$ and $U_{l_i''}$:

$$\left(\bigcup_{j=1}^{k_T} T_{l_{i_j}'}\right) \cup \left(\bigcup_{j=1}^{k_U} U_{l_{i_j}''}\right) = A,\ k_U \geq 1,\ k_T + k_U < n.$$

$$(\forall i_1,\ i_2)\ T_{l_{i_1}'} \cap T_{l_{i_2}'} = \varnothing,\ U_{l_{i_1}''} \cap U_{l_{i_2}''} = \varnothing,$$

$$\left(\forall\ (a_i, a_i) \in T_{l_{i_1}'}\right) a_i \succ a_j \vee a_j \succ a_i,\ \left(\forall\ (a_i, a_i) \in U_{l_{i_1}''}\right) a_i \sim a_j,$$

$$\left(\forall\ a_i \in T_{l_{i_1}'}\right) \& \left(\forall\ a_j \in T_{l_{i_2}'}\right)\ if\ l_{i_1}' > l_{i_2}',\ then\ a_i \succ a_j,$$

$$\left(\forall\ a_i \in T_{l_{i_1}'}\right) \& \left(\forall\ a_j \in U_{l_{i_2}''}\right)\ if\ l_{i_1}' > l_{i_2}'',\ then\ a_i \succ a_j,$$

$$\left(\forall\ a_i \in U_{l_{i_1}''}\right) \& \left(\forall\ a_j \in U_{l_{i_2}''}\right)\ if\ l_{i_1}'' > l_{i_2}'',\ then\ a_i \succ a_j.$$

Define the preference incompatibility assessment (distance) function:

$$d(\succcurlyeq^{(1)}, \succcurlyeq^{(2)}) = \left|\succcurlyeq^{(1)} \setminus I_A \setminus \succcurlyeq^{(2)}\right| + \left|\succcurlyeq^{(2)} \setminus I_A \setminus \succcurlyeq^{(1)}\right|$$

Here $I_A = \{(a_1, a_1),\ (a_2, a_2), \ldots, (a_n, a_n)\}$ is set A identity relation, $|\ldots|-$ the number of elements in the set . The relation S can be expressed in a matrix

$$M_S = \|m_{ij}\|_{n \times n}\ ,\ m_{ij} = \begin{cases} 1, \text{ kai } (a_i, a_j) \in S, \\ 0, \text{ kai } (a_i, a_j) \notin S. \end{cases}$$

If the relations $\succcurlyeq^{(1)}$ and $\succcurlyeq^{(2)}$ are expressed as matrices $M_{\succcurlyeq^{(1)}} = \|a_{ij}^{(1)}\|_{n \times n}$ and $M_{\succcurlyeq^{(2)}} = \|a_{ij}^{(2)}\|_{n \times n}$, then

$$d(\succcurlyeq^{(1)}, \succcurlyeq^{(2)}) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j \\ a_{ij}^{(1)} \neq a_{ij}^{(2)}}}^{n} a_{ij}^{(1)} + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j \\ a_{ij}^{(2)} \neq a_{ij}^{(1)}}}^{n} a_{ij}^{(2)}.$$

Example. This example illustrates how the distance between two preferences is calculated.

Suppose we have the set $A = \{1, 2, \ldots, 11\}$ and preference $\succcurlyeq$:

$$(3) \sim (5) \succ (7) \sim (9) \succ (1) \succ (2) \sim (4) \succ (6) \sim (11) \succ (10) \succ (8). \quad (1)$$

Write the corresponding blocks of the set A:

$$U_1 = \{a_3, a_5\},\ U_2 = \{a_7, a_9\},\ T_3 = \{a_1\},\ U_4 = \{a_2, a_4\},\ U_5 = \{a_6, a_{11}\},\ T_6 = \{a_{10}, a_8\},\ a_{10} \succ a_8.$$

Let us compare preference $\succcurlyeq$ with the preference $\succ$:

$$(3) \succ (5) \succ (6) \succ (9) \succ (4) \succ (7) \succ (1) \succ (2) \succ (11) \succ (8) \succ (10). \quad (2)$$

Matrices of the samples under consideration:

$$A_{\succcurlyeq} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$A_{\succ} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

$$A_{\succ \setminus I_A \setminus \succcurlyeq} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_{\succcurlyeq\setminus I_A\setminus\succ} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The value of the distance function

$$d(\succcurlyeq,\succ) = d(\succ,\succcurlyeq) = \left|A_{\succ\setminus I_A\setminus\succcurlyeq}\right| + \left|A_{\succcurlyeq\setminus I_A\setminus\succ}\right| = 8 + 12 = 20.$$

This is the measure of the preferences (1) and (2) incompatibility.

Remark. The preference $\succcurlyeq$ in this example can be expressed by giving the scores for set $A = \{1,\ 2, \ldots, 11\}$ items

$$\succcurlyeq = (6,5,8,5,8,4,7,2,7,3,4).$$

The same preference can also be expressed in this way:

$$\succcurlyeq = (5,4,10,4,10,3,9,1,9,2,3).$$

We express preference $\succ$ as follows:

$$\succ = (7,8,1,5,2,3,6,10,4,11,9)$$

Note that preference $\succ$ is expressed unambiguously.

3. Description of Life Goals

The basic human motivational factors identified as general worldview values capture both real and desired personal traits [23], which determinate life goals. Major life goals encompass goals that influence personality development in long-term perspectives.

The taxonomy of major life goals used in this study is close to the life goals of college students [24], which is grounded by seven domains of life goals: Aesthetic, Economic, Family/Relationship, Hedonistic, Political, Religious, and Social. To more accurately describe the evaluation criteria, Table 1 provides an expanded list of life goals [25].

The applied survey data processing methodologies did not provide an assessment of the reliability of the respondents' answers. To date, there is no approved methodology to assess the reliability of sociological research data. It is not clear what motives were followed and in what state the respondents conducted the sociological survey, and how they understood the questions and tasks presented. It is also not clear how honestly the respondents conducted the survey. It is important to develop methodologies for filtering data from sociological surveys. Tools based on these methodologies will help assess the reliability of sociological survey data in the future. Studies of student life goal assessment have already been conducted without applying data reliability filtering methodologies [26,27]. However, the reliability of these study data is unclear.

Table 1. Evaluation of life goals factors.

No.	Life Goals
x1	Business/Career/Studies;-Usually the key segment in our lives. Business is for entrepreneurs, career for employees and studies for students
x2	Finance/Wealth; -How rich you are. The amount of wealth, assets, material possessions you have
x3	Health/Fitness;Your state of health as well as your lifestyle. Diet, sleep, exercise falls here
x4	Social/Friends;How you are faring in your social circle
x5	Family-Your parents, siblings, next of kin, relatives, or even your guardians
x6	Love; the amount of love you feel in your life. While it can represent the status of your relationship with your spouse/boyfriend/girlfriend, it does not have to be the case. Love here does not refer to romantic love but is about universal love
x7	Recreation/Fun;Your recreation and enjoyment in life
x8	Contribution;How you are giving back to the society. Social cause. Humanitarian activities.
x9	Personal growth;Your personal development as a whole
x10	Spiritual;Your connection with the universe. Some call it higher power/God/higher self
x11	Self-image-How you see yourself
The students evaluate the criteria (life goals) according to the following rules: Ranking (R): Rank all the general life goal criteria in your life from 1 to 11 (1-the least important; 11-the most important). Scores now (N) Rate your satisfaction life goal criteria level today in each of the every "slices" of your life goals above, using the following scale: from 1 = Totally Dissatisfied to 10 = Totally Satisfied. Scores in the future (F): Rate your satisfaction life goal criteria level in the future in each of the every "slices" of your life goals above, using the following scale: from 1 = Totally Dissatisfied to 10 = Totally Satisfied.	

The experiment participants, 19–22 years old, second and third-year students, were randomly selected from different faculties of VilniusTech: Vilnius Gediminas Technical University (Antanas Gustaitis Aviation Institute, Faculty of Environmental Engineering, Faculty of Architecture, Faculty of Business and Management, Faculty of Civil Engineering, Faculty of Creative Industries, Faculty of Electronics, Faculty of Fundamental Sciences, Faculty of Mechanics, Faculty of Transport Engineering), and Wuhan Sports University, Wuhan, China (Physical Education College, Chinese Martial Arts College, College of Health Science, Sports Journalism and Foreign Language Faculty).

Evaluation criteria: the selected students evaluated 11 life goals factors (criteria) according to Chua [25] (Table 1) in two ways-by ranking and by assigning scores (10-point scale).

First, 11 values x1–x11 had to be ranked by importance (1-the most important, 2-in the second place, and so on), and students had to make a preference $\succ$. The same students were also asked to score the same 11 values from 1 to 10, i.e., to make a preference $\succeq \bigcap$. Note that the preferences $\succ$ and $\succeq \bigcap$ constructed in this way cannot coincide, and the minimum value of the function d ($\succ$,$\succeq \bigcap$) is equal to 1. The maximum value of this function indicates a complete mismatch of preferences and is equal to 110 = n^2-n, when n = 11. In Tables 2 and 3, the preferences of Lithuanian women and men, respectively, distances between ranks (R) and scores now (N) dRN, ranks (R) and scores in the future (F) dRF are presented for the first dataset of Lithuanian students (51 women and 106 men). Values of distance functions dRN and dRF were calculated by applying formulas presented in Section 2.

Table 2. Ranks, scores assigned to 11 values by 51 women now and in the future and values of the distance functions dRN, dRF for Lithuanian data set 1.

No.	Ranks (R)	Scores Now (N)	Scores in the Future (F)	dRN	dRF
1	9; 8; 3; 2; 4; 6; 5; 10; 1; 11; 7	10; 5; 6; 10; 7; 8; 9; 2; 4; 1; 3	9; 10; 10; 6; 7; 8; 1; 10; 5; 3; 4	39	63
2	1; 2; 3; 6; 5; 4; 9; 10; 7; 11; 8	10; 5; 8; 6; 9; 10; 4; 6; 5; 1; 7	10; 5; 8; 4; 9; 10; 4; 8; 7; 7; 8	27	40
3	7; 5; 1; 6; 3; 8; 9; 10; 2; 11; 4	6; 6; 8; 7; 9; 7; 8; 5; 8; 7; 8	10; 10; 10; 10; 10; 10; 10; 10; 10; 10; 10	34	55
⋮	⋮	⋮	⋮	⋮	⋮
49	6; 9; 3; 4; 1; 8; 5; 11; 2; 7; 10	9; 8; 8; 10; 10; 5; 7; 2; 9; 8; 6	10; 10; 9; 10; 10; 10; 8; 7; 9; 9; 8	21	42
50	4; 8; 5; 7; 1; 2; 9; 10; 3; 11; 6	8; 6; 8; 8; 8; 5; 8; 4; 6; 5; 7	9; 9; 10; 8; 10; 8; 8; 8; 10; 6; 9	44	26
51	3; 8; 5; 4; 1; 2; 9; 10; 7; 11; 6	7; 7; 8; 6; 10; 10; 8; 4; 7; 3; 6	10; 8; 9; 8; 10; 10; 7; 7; 8; 6; 8	28	12

Table 3. Ranks, scores assigned to 11 values by 106 men now and in the future and values of the distance functions dRN, dRF for Lithuanian data set 1.

No.	Ranks (R)	Scores Now (N)	Scores in the Future (F)	dRN	dRF
1	3; 1; 2; 5; 11; 6; 9; 4; 7; 8; 10	2; 1; 3; 4; 7; 5; 9; 10; 10; 6; 8	9; 2; 6; 3; 5; 7; 10; 8; 6; 8; 10	81	71
2	1; 5; 3; 4; 2; 6; 7; 10; 8; 11; 9	10; 6; 9; 9; 10; 9; 7; 3; 10; 1; 10	10; 10; 10; 9; 10; 10; 10; 8; 10; 1; 10	33	38
3	8; 10; 3; 4; 5; 7; 2; 9; 1; 11; 6	8; 8; 6; 7; 7; 4; 7; 7; 5; 10; 7	8; 9; 9; 9; 8; 8; 7; 7; 10; 10; 8	83	53
⋮	⋮	⋮	⋮	⋮	⋮
104	8; 7; 4; 6; 1; 3; 5; 11; 2; 10; 9	8; 3; 6; 5; 10; 7; 3; 2; 9; 4; 6	7; 8; 10; 10; 10; 10; 8; 5; 10; 7; 8	24	18
105	6; 5; 2; 7; 1; 3; 9; 10; 4; 11; 8	7; 6; 5; 9; 10; 3; 10; 2; 8; 1; 4	10; 9; 10; 8; 6; 7; 4; 8; 3; 9; 5	41	55
106	5; 3; 4; 8; 7; 6; 9; 10; 2; 11; 1	5; 2; 5; 8; 10; 1; 3; 1; 5; 7; 8	10; 8; 7; 7; 10; 10; 5; 5; 5; 5; 10	49	37

4. Data Filtering and Life Goals Prioritisation Results

4.1. Description of the Data Filtering Procedure

From Tables 2 and 3, we can see that some students did not notice that the score scale ranged from 1 to 10 and used 11 points. In this case, the value of 11 has being changed to 10. Also, the value of 0 has being changed to 1. Data errors could have occurred both due to elementary negligence as well as due to a completely irresponsible approach to the survey. For example, using 11 points, it is possible to mechanically repeat the preference $\succ$ and write $\succcurlyeq=\succ$. Therefore, not only high but also very low d values can be suspicious. However, the real d values in Tables 2 and 3 show that there are practically no very small d values. Notice that the assessments in the column Ranks (R) refer to the student's view of the values themselves, i.e., how they are assessed "objectively". Evaluations in the column Scores now (N) are obtained when the student evaluates how significant this value is to him or her, i.e., the same values are assessed "subjectively". Subjective assessments were obtained when a student was asked to score the same values now and, in the future, i.e., how he or she imagines the dynamics of his values, how they can change in his or her value system later. For various reasons, these assessments cannot coincide completely; however,

they should be "sufficiently matched". The aim of our study was to construct a criterion that would allow to distinguish unreliable data, i.e., eliminate questionable rows from the table and then construct generalized student opinion preferences.

Large values of the function d indicate poor compatibility between the two preferences, and such survey results should be rejected as unreliable. Thus, we can apply a filter:

$$d(\succcurlyeq, \succ) < d_0.$$

To determine the value of the filter threshold d_0, we will generate random preferences and evaluate the values of the function d. For real data, the null hypothesis of a normal distribution of the values of the function d was tested. The hypothesis was not rejected. So, 10,000 random normally distributed preferences $\succcurlyeq$ and $\succ$ were generated 10 times. Table 4 shows calculated means and standard deviations of the function d for randomly generated data.

Table 4. Means ($\overline{d}$) and standard deviations (σ) of the function d for randomly generated data.

$\overline{d}$	54.84	55.1	54.89	55.01	55.14	54.93	54.93	55.12	55.16	54.95
σ	12.75	12.57	12.58	12.5	12.69	12.5	12.55	12.44	12.67	12.51

Assuming a normal distribution, 83% of the values of the function d belong to the range $[\overline{d} - \sigma, \overline{d} + 3\sigma]$. We have to reject randomly generated values as unreliable data. Thus, we can consider only the data satisfying inequality

$$d(\succcurlyeq, \succ) < d_0 = \overline{d} - \sigma \approx 42.$$

The values of function d greater than $d_0 = 42$ indicate that corresponding data are unreliable or random. Both ranks (R) and scores now (N) and ranks (R) and scores in the future (F) must be well matched. The measures of compatibility are the corresponding distances $d(R, N)$ and $d(R, F)$. Table 5 shows filtered data when $d(R, N) < d_0$ and $d(R, F) < d_0$ for women and men, respectively.

Table 5. Filtered data for women and men (Lithuania, data set 1) and corresponding $d(R, N)$, $d(R, F)$ values.

Women			**Men**					
No.	$d(R,N)$	$d(R,F)$	**No.**	$d(R,N)$	$d(R,F)$	**No.**	$d(R,N)$	$d(R,F)$
1 (2)	27	40	1 (2)	33	38	31 (52)	24	24
2 (5)	34	26	2 (4)	26	24	32 (57)	31	29
3 (6)	31	33	3 (5)	18	15	33 (58)	34	25
4 (7)	28	23	4 (7)	32	28	34 (60)	30	14
5 (9)	28	23	5 (14)	15	17	35 (61)	21	16
6 (10)	24	33	6 (15)	21	36	36 (63)	28	28
7 (12)	35	33	7 (17)	20	13	37 (64)	40	33
8 (15)	32	37	8 (18)	39	38	38 (67)	22	16
9 (16)	23	38	9 (19)	35	32	39 (68)	18	18
10 (18)	12	17	10 (23)	37	13	40 (70)	26	22
11 (20)	34	36	11 (24)	30	37	41 (71)	9	41
12 (21)	31	38	12 (25)	27	36	42 (78)	18	18
13 (22)	37	33	13 (26)	11	24	43 (80)	22	22
14 (23)	26	40	14 (27)	15	29	44 (81)	35	38

Table 5. *Cont.*

Women			Men					
No.	$d(R,N)$	$d(R,F)$	**No.**	$d(R,N)$	$d(R,F)$	**No.**	$d(R,N)$	$d(R,F)$
15 (26)	14	29	15 (30)	33	36	45 (83)	28	25
16 (28)	14	28	16 (31)	36	21	46 (84)	34	25
17 (29)	22	34	17 (32)	26	15	47 (85)	29	25
18 (30)	27	36	18 (34)	26	38	48 (90)	34	26
19 (31)	18	31	19 (37)	20	13	49 (92)	25	36
20 (34)	17	36	20 (38)	20	13	50 (98)	29	38
21 (35)	36	24	21 (39)	34	22	51 (99)	38	23
22 (36)	33	37	22 (40)	32	20	52 (100)	27	37
23 (37)	31	35	23 (41)	17	14	53 (102)	35	38
24 (40)	22	39	24 (42)	37	18	54 (103)	27	24
25 (42)	35	26	25 (43)	24	31	55 (104)	24	18
26 (43)	24	29	26 (46)	38	38			
27 (45)	11	31	27 (48)	41	31			
28 (47)	15	14	28 (49)	21	16			
29 (48)	28	33	29 (50)	34	28			
30 (51)	28	12	30 (51)	25	20			

Numbers given in parentheses in Table 5 correspond to the numbers of respondents in Tables 2 and 3, where initial data for Lithuanian data set 1 are proposed. We see that from 51 women preferences, 30 (58.8%) were selected, and from 106 men preferences, 55 (51.9%) were selected, according to the filter with threshold $d_0 = 42$.

4.2. Prioritisation Results for Lithuanian Data Set 1

Table 6 presents basic statistics for Lithuanians women and men filtered data.

Table 6. Minimum, maximum values, averages, and standard deviations of d (R, N), d (R, F) for women and men filtered data (Lithuania, data set 1).

	Women			
	Min	Max	Average	Standard deviation
$d(R,N)$	11	37	25.9	7.68
$d(R,F)$	12	40	30.8	7.44
	Men			
	Min	Max	Average	Standard deviation
$d(R,N)$	9	41	27.47	7.67
$d(R,F)$	13	41	25.69	8.66

All values presented in Table 6 satisfy the inequality $x_i \geq \bar{x} - 3\sigma$ and almost all values satisfy the inequality $x_i \geq \bar{x} - 2\sigma$, i.e., there are no very small values. Statistical data for women and men do not differ significantly. For each of 11 values described in Table 1, we calculate the average ranks (R_j) and the average scores assigned now (N_j) and in the future (F_j) by the formulas below:

$$R_j = \frac{1}{n_{stud}} \sum_{i=1}^{n_{stud}} \left(11 - R_j^{(i)}\right),\ j = 1, 2, \ldots, 11,$$

$$N_j = \frac{1}{n_{stud}} \sum_{i=1}^{n_{stud}} N_j^{(i)},\ j = 1, 2, \ldots, 11,$$

$$F_j = \frac{1}{n_{stud}} \sum_{i=1}^{n_{stud}} F_j^{(i)},\ j = 1, 2, \ldots, 11.$$

The results of the calculations are presented in Table 7.

Table 7. Averages of 11 values ranks (*R_j*), average scores assigned now (*N_j*) and in the future (*F_j*) for women and men filtered data (Lithuania, data set 1).

Women ($n_{stud} = 30$)											
Values	1	2	3	4	5	6	7	8	9	10	11
R_j	6.13	4.27	7.37	6.00	8.87	7.77	2.93	1.17	5.77	1.90	2.83
N_j	7.77	6.30	7.90	7.10	9.20	8.60	6.87	4.13	7.60	4.30	6.90
F_j	9.17	8.60	9.50	8.30	9.67	9.47	8.00	7.37	9.00	6.50	8.57
Men ($n_{stud} = 55$)											
Values	1	2	3	4	5	6	7	8	9	10	11
R_j	6.31	5.44	7.49	6.02	8.05	7.22	3.56	1.45	6	0.84	2.62
N_j	7.20	6.05	7.27	7.42	8.51	7.45	6.62	3.47	7.09	3.13	6.42
F_j	9.29	9.07	9.4	8.73	9.38	9.38	8.13	6.4	8.89	4.93	8.13

Based on these data, we construct preferences in two ways: we ranked the data in Table 7 and received the strict ($\succ$) preferences ($\boldsymbol{R}, \boldsymbol{N}, \boldsymbol{F}$). Then we round the numbers in Table 7 to the integer, calculated the average scores and received the corresponding unstrict ($\succcurlyeq$) preferences ($\overline{\boldsymbol{R}}, \overline{\boldsymbol{N}}, \overline{\boldsymbol{F}}$).

Results for women:

$$R = (4, 7, 3, 5, 1, 2, 8, 11, 6, 10, 9)$$

$$N = (4, 9, 3, 6, 1, 2, 8, 11, 5, 10, 7),$$

$$F = (4, 6, 2, 8, 1, 3, 9, 10, 5, 11, 7),$$

$$\overline{R} = (6, 4, 7, 6, 9, 8, 3, 1, 6, 2, 3)$$

$$\overline{N} = (8, 6, 8, 7, 9, 9, 7, 4, 8, 4, 7)$$

$$\overline{F} = (9, 9, 10, 8, 10, 9, 8, 7, 9, 7, 9).$$

Results for men:

$$R = (4, 7, 2, 5, 1, 3, 8, 10, 6, 11, 9)$$

$$N = (5, 9, 4, 3, 1, 2, 7, 10, 6, 11, 8),$$

$$F = (4, 5, 1, 7, 2, 3, 8, 10, 6, 11, 9),$$

$$\overline{R} = (6, 5, 7, 6, 8, 7, 4, 1, 6, 1, 3)$$

$$\overline{N} = (7, 6, 7, 7, 9, 7, 7, 3, 7, 3, 6)$$

$$\overline{F} = (9, 9, 9, 9, 9, 9, 8, 6, 9, 5, 8).$$

We will present the constructed preferences in a more convenient form. These results show high (green colour) and low (grey colour) ranked life goals for women and men (Lithuania, data set 1).

Women:

$$\boldsymbol{R} : (5) \succ (6) \succ (3) \succ (1) \succ (4) \succ (9) \succ (2) \succ (7) \succ (11) \succ (10) \succ (8),$$

$$N : (5) \succ (6) \succ (3) \succ (1) \succ (4) \succ (9) \succ (11) \succ (7) \succ (2) \succ (10) \succ (8),$$
$$F : (5) \succ (3) \succ (6) \succ (1) \succ (9) \succ (2) \succ (11) \succ (4) \succ (7) \succ (8) \succ (10),$$
$$\overline{R} : (5) \succ (6) \succ (3) \succ (1) \sim (4) \sim (9) \succ (2) \succ (7) \sim (11) \succ (10) \succ (8)$$
$$\overline{N} : (5) \sim (6) \succ (1) \sim (3) \sim (9) \succ (4) \sim (7) \sim (11) \succ (2) \succ (10) \sim (8)$$
$$\overline{F} : (5) \sim (3) \succ (1) \sim (2) \sim (6) \sim (9) \sim (11) \succ (4) \sim (7) \succ (10) \sim (8).$$

Men:

$$R : (5) \succ (3) \succ (6) \succ (1) \succ (4) \succ (9) \succ (2) \succ (7) \succ (11) \succ (8) \succ (10)$$
$$N : (5) \succ (6) \succ (4) \succ (3) \succ (1) \succ (9) \succ (7) \succ (11) \succ (2) \succ (8) \succ (10),$$
$$F : (3) \succ (5) \succ (6) \succ (1) \succ (2) \succ (9) \succ (4) \succ (7) \succ (11) \succ (8) \succ (10),$$
$$\overline{R} : (5) \succ (3) \sim (6) \succ (1) \sim (4) \sim (9) \succ (2) \succ (7) \succ (11) \succ (8) \sim (10)$$
$$\overline{N} : (5) \succ (1) \sim (3) \sim (4) \sim (6) \sim (7) \sim (9) \succ (2) \sim (11) \succ (8) \sim (10)$$
$$\overline{F} : (1) \sim (2) \sim (3) \sim (4) \sim (5) \sim (6) \sim (9) \succ (7) \sim (11) \succ (8) \succ (10).$$

The following values were highly rated by both Lithuanian women and men: x5-family, x6-love, x3-health. Meanwhile, the lowest-rated values by both Lithuanian women and men are x8-contribution (humanitarian activities) and x10-spiritual, with x11-self-image being lower rated by men than by women. Women rated values x2-finance and x7-recreation/fun lower than men. However, x1-business, x2-finance, and x9-personal growth were higher evaluated by both women and men in the future.

4.3. Prioritisation Results for Other Datasets

The second set of students from Lithuania included 128 women and 810 men; after the filtering procedure described in the previous section, 77 (60%) women and 461 (57%) men remained. Table 8 presents averages of 11 values ranked for women and men and average scores assigned now and in the future.

Table 8. Averages of 11 values ranks ($\boldsymbol{R}_j$), average scores assigned now ($\boldsymbol{N}_j$) and in the future ($\boldsymbol{F}_j$) for women and men filtered data (Lithuania, data set 2).

Women ($n_{stud} = 77$)											
Values	1	2	3	4	5	6	7	8	9	10	11
R_j	6.1	4.13	7.6	5.79	9.16	7.51	3.03	1.22	5.7	1.34	3.43
N_j	7.87	6.32	7.77	7.58	9.25	8.34	7.01	4.31	7.47	4.38	7.57
F_j	9.47	9.08	9.64	8.87	9.86	9.68	8.45	7.21	9.23	6.29	9.04
Men ($n_{stud} = 461$)											
Values	1	2	3	4	5	6	7	8	9	10	11
R_j	6.51	5.23	7.15	5.85	8.8	7.03	3.54	1.28	5.62	0.9	3.09
N_j	7.1	6.21	7.46	7.81	8.88	7.4	6.8	3.72	6.94	3.56	6.8
F_j	9.3	9.03	9.28	8.8	9.65	9.39	8.01	6.3	8.92	5.18	8.29

The constructed preferences for women:

$$R : (5) \succ (3) \succ (6) \succ (1) \succ (4) \succ (9) \succ (2) \succ (11) \succ (7) \succ (10) \succ (8),$$
$$N : (5) \succ (6) \succ (1) \succ (3) \succ (4) \succ (11) \succ (9) \succ (7) \succ (2) \succ (10) \succ (8),$$
$$F : (5) \succ (6) \succ (3) \succ (1) \succ (9) \succ (2) \succ (11) \succ (4) \succ (7) \succ (8) \succ (10),$$

$$\overline{R}: (5) \succ (6) \sim (3) \succ (1) \sim (4) \sim (9) \succ (2) \succ (7) \sim (11) \succ (8) \sim (10)$$
$$\overline{N}: (5) \succ (1) \sim (3) \sim (4) \sim (6) \sim (11) \succ (7) \sim (9) \succ (2) \succ (8) \sim (10)$$
$$\overline{F}: (3) \sim (5) \sim (6) \succ (1) \sim (2) \sim (4) \sim (9) \sim (11) \succ (7) \succ (8) \succ (10).$$

Preferences for men:

$$R: (5) \succ (3) \succ (6) \succ (1) \succ (4) \succ (9) \succ (2) \succ (7) \succ (11) \succ (8) \succ (10)$$
$$N: (5) \succ (4) \succ (3) \succ (6) \succ (1) \succ (9) \succ (11) \succ (7) \succ (2) \succ (8) \succ (10),$$
$$F: (5) \succ (6) \succ (1) \succ (3) \succ (2) \succ (9) \succ (4) \succ (11) \succ (7) \succ (8) \succ (10),$$
$$\overline{R}: (5) \succ (1) \sim (3) \sim (6) \succ (4) \sim (9) \succ (2) \succ (7) \succ (11) \succ (8) \sim (10)$$
$$\overline{N}: (5) \succ (4) \succ (1) \sim (3) \sim (6) \sim (7) \sim (9) \sim (11) \succ (2) \succ (8) \sim (10)$$
$$\overline{F}: (5) \succ (1) \sim (2) \sim (3) \sim (4) \sim (6) \sim (9) \succ (7) \sim (11) \succ (8) \succ (10).$$

These values were highly ranked by both Lithuanian women and men: x5-family, x6-love (more important for women), x3-health, and x1-business, while x4-social/friends was more highly rated by men than by women. The lowest rated by both men and women were x8-contribution, x10-spiritual. x11-self-image (less important for men), x7-recreation/fun and x2-finance. The results are very similar to the results for the first data set of Lithuanian students.

The third data set is from the Faculty of Physical Education and English at Wuhan Sports University. The data set includes data of 101 women and 120 men. After the filtering procedure, 33 (33%) women and 24 (20%) men were selected for further data processing. The averages of the values used to determine the group priorities of 11 values are presented in Table 9. Thus, one can notice that the China student set filtered at least twice as much data as the Lithuanian student sets (20–33% left in China data set *versus* 51–60% left in Lithuanian data sets).

Table 9. Averages of 11 values ranks ($\boldsymbol{R}_j$), average scores assigned now ($\boldsymbol{N}_j$) and in the future ($\boldsymbol{F}_j$) for women and men filtered data (China, data set 1).

Women ($n_{stud} = 33$)											
Values	1	2	3	4	5	6	7	8	9	10	11
R_j	5.79	3.55	7.06	5.55	9.06	6.58	1.82	1.97	5.88	4.15	3.61
N_j	5.67	3.52	5.88	6.73	8.67	7.73	5.64	3.94	5.64	5.03	5.76
F_j	8.36	7.64	8.39	8.24	9.24	9.12	7.30	6.88	8.09	7.18	7.67
Men ($n_{stud} = 24$)											
Values	1	2	3	4	5	6	7	8	9	10	11
R_j	7.38	4.17	7.58	6.17	8.83	6.5	1.75	2.54	4.92	2.5	2.67
N_j	5.96	3.58	7.21	6.38	8.25	6.13	4.67	4.08	5.54	4.29	5.83
F_j	8.71	8.29	9.29	8.58	9.21	8.58	6.67	7.13	8.79	7.13	8.00

So, we can calculate group preferences for women:

$$R: (5) \succ (3) \succ (6) \succ (9) \succ (1) \succ (4) \succ (10) \succ (11) \succ (2) \succ (8) \succ (7),$$
$$N: (5) \succ (6) \succ (4) \succ (3) \succ (11) \succ (1) \succ (7) \succ (9) \succ (10) \succ (8) \succ (2),$$
$$F: (5) \succ (6) \succ (3) \succ (1) \succ (4) \succ (9) \succ (11) \succ (2) \succ (7) \succ (10) \succ (8),$$
$$\overline{R}: (5) \succ (3) \sim (6) \succ (1) \sim (4) \sim (9) \succ (2) \sim (10) \sim (11) \succ (7) \sim (8)$$

$$\overline{N} : (5) \succ (6) \sim (4) \succ (1) \sim (3) \sim (7) \sim (9) \sim (11) \succ (10) \succ (2) \sim (8)$$
$$\overline{F} : (5) \sim (6) \succ (1) \sim (2) \sim (3) \sim (4) \sim (9) \sim (11) \succ (7) \sim (8) \sim (10)$$

and preferences for men:

$$R : (5) \succ (3) \succ (1) \succ (6) \succ (4) \succ (9) \succ (2) \succ (11) \succ (8) \succ (10) \succ (7)$$
$$N : (5) \succ (3) \succ (4) \succ (6) \succ (1) \succ (11) \succ (9) \succ (7) \succ (10) \succ (8) \succ (2),$$
$$F : (3) \succ (5) \succ (9) \succ (1) \succ (4) \succ (6) \succ (2) \succ (11) \succ (8) \succ (10) \succ (7),$$
$$\overline{R} : (5) \succ (3) \succ (1) \sim (6) \succ (4) \succ (9) \succ (2) \succ (8) \sim (10) \sim (11) \succ (7)$$
$$\overline{N} : (5) \succ (3) \succ (1) \sim (4) \sim (6) \sim (9) \sim (11) \succ (7) \succ (2) \sim (8) \sim (10)$$
$$\overline{F} : (1) \sim (3) \sim (4) \sim (5) \sim (6) \sim (9) \succ (2) \sim (11) \succ (7) \sim (8) \sim (10).$$

The values highly estimated by China students are x5-family, x6-love (more important for women), x3-health (more important for men), x1-business (more important for men), x4-social/friends. The lowest rated by both China men and women were x8-contribution (less important for women), x7-recreation/fun (less important for men), x10-spiritual and x2-finance.

The x11-self-image value is more important for China students than for Lithuanians. We also noticed that x7-recreation/fun is less important for China students than for Lithuanians. Lithuanian students of both sexes are remarkably united in determining the values with the lowest priority, i.e., x8 and x10.

The fourth data set is from the Physical Education College, Chinese Martial Arts College, College of Health Science, Sports Journalism and Foreign Language Faculty at Wuhan Sports University. The data set includes 149 women and 169 men; after the filtering procedure, 37 (25%) women and 40 (24%) men remained. Table 10 presents averages for 11 values ranks, average scores assigned now and in the future.

Table 10. Averages of 11 values ranks (*R_j*), average scores assigned now (*N_j*) and in the future (*F_j*) for women and men filtered data (China, data set 2).

Women ($n_{stud} = 37$)											
Values	1	2	3	4	5	6	7	8	9	10	11
R_j	6.70	4.16	7.14	4.38	9.59	7.16	1.73	2.22	5.81	3.11	3.00
N_j	6.03	4.03	6.32	6.05	8.95	7.73	5.73	4.03	5.89	4.86	5.86
F_j	8.14	7.62	8.65	8.03	9.59	8.89	6.54	7.08	7.78	6.73	7.38
Men ($n_{stud} = 40$)											
Values	1	2	3	4	5	6	7	8	9	10	11
R_j	6.43	3.23	6.48	5.50	9.23	6.80	2.00	3.03	6.00	3.50	2.83
N_j	6.20	3.18	7.13	6.73	8.88	7.48	5.33	4.78	6.15	5.15	5.83
F_j	8.53	7.60	8.60	8.43	9.53	8.98	7.35	7.58	8.08	7.23	7.58

Calculate preferences for women:

$$R : (5) \succ (6) \succ (3) \succ (1) \succ (9) \succ (4) \succ (2) \succ (10) \succ (11) \succ (8) \succ (7),$$
$$N : (5) \succ (6) \succ (3) \succ (4) \succ (1) \succ (9) \succ (11) \succ (7) \succ (10) \succ (2) \succ (8),$$
$$F : (5) \succ (6) \succ (3) \succ (1) \succ (4) \succ (9) \succ (2) \succ (11) \succ (8) \succ (10) \succ (7),$$
$$\overline{R} : (5) \succ (1) \sim (3) \sim (6) \succ (9) \succ (2) \sim (4) \succ (10) \sim (11) \succ (7) \sim (8)$$
$$\overline{N} : (5) \succ (6) \succ (1) \sim (3) \sim (4) \sim (7) \sim (9) \sim (11) \succ (10) \succ (2) \sim (8)$$

$$\overline{F}: (5) \succ (3) \sim (6) \succ (1) \sim (2) \sim (4) \sim (9) \succ (7) \sim (8) \sim (10) \sim (11)$$

and preferences for men:

$$R: (5) \succ (6) \succ (3) \succ (1) \succ (9) \succ (4) \succ (10) \succ (2) \succ (8) \succ (11) \succ (7)$$

$$N: (5) \succ (6) \succ (3) \succ (4) \succ (1) \succ (9) \succ (11) \succ (7) \succ (10) \succ (8) \succ (2),$$

$$F: (5) \succ (6) \succ (3) \succ (1) \succ (4) \succ (9) \succ (2) \succ (8) \succ (11) \succ (7) \succ (10),$$

$$\overline{R}: (5) \succ (6) \succ (1) \sim (3) \sim (4) \sim (9) \succ (10) \succ (2) \sim (8) \sim (11) \succ (7)$$

$$\overline{N}: (5) \succ (3) \sim (4) \sim (6) \succ (1) \sim (9) \sim (11) \succ (7) \sim (8) \sim (10) \succ (2)$$

$$\overline{F}: (5) \succ (1) \sim (3) \sim (6) \succ (2) \sim (4) \succ (8) \sim (9) \sim (11) \succ (7) \sim (10).$$

Note that priority $\succ$ for all students of the fourth data set is identical: x5 $\succ$ x6 $\succ$ x3 (x5-family, x6-love, x3-health). x1-business is also highly rated. The lowest rated are x8-contribution (less important for women), x10-spiritual, x2-finance, and x11-self-image are less important for Chinese men while x7-recreation/fun is less important for all Chinese students.

5. Discussion

The social sciences are based on human behavioural research data. Generalizations and important social decisions are made based on these data. However, these studies still lack procedures and tools for assessing data reliability on human behaviour. Such tools would make it possible to make reasonable generalizations that globally reflect the characteristics of the population being studied. Failure to assess the reliability of sociological data harms the quality of research [28]. The assessment of data reliability should become a mandatory procedure in the practice of sociological research. The solving of this problem requires reliable tools.

This study aimed to identify the reliability criteria of sociological data and develop a data filtering tool to help exclude elements that cause information noise. This allows researchers to obtain reliably similar results in a replicated study. The reliability of research data in sociological research can be summarised by saying that what is right for one representative of the target group will be the same or similar for all members of the target group [29].

The example of sociological research presented in this article allowed us to avoid substantial sociological differences in the analysed data. The provided combination of data processing and filtering tools can be adapted to analyse various data types and volumes. It can effectively identify data that creates information noise which leads to misinterpretation and erroneous conclusions and decisions.

The digitization and processing of large-scale information from target audiences should be consistent. The consistency is achieved by the following means:

1. Proper questionnaire design;
2. Information gathering;
3. Calculation of distances between preferences;
4. Setting a threshold for data filtering;
5. Selection of reliable data using ascertained threshold;
6. Calculation of average ranks and average estimates;
7. Calculation of group preferences.

The above measures can help decision-making in many areas and ensure sustainable development. The model of differentiation between reliable and unreliable information proposed in this study can help better understand the regularities of segmentation and differentiation of information. The differentiation of sociological research data into reliable and unreliable information can be applied to deficiencies in respondents' basic knowledge, skills, and metacompetences necessary for selection or assessment, misinterpretations of the

information submitted for the assessment, insufficient effort and superficial assessment of the information. The differentiation between reliable and unreliable information analysed by the technologies that generate data can help identify errors in the system process.

The development of data reliability assessment and filtering tools should be targeted by addressing specific challenges. However, the tools used to assess the reliability of the data should meet the critical criteria [30]. To achieve data reliability, a good index of reliability should have the following properties:

1. It should assess the agreement between two or more observers who describe each unit of the analysis independently of each other. For more than two observers, this measure should be (a) independent of the number of observers employed and (b) invariant to the permutation and selective participation of the observers.
2. It should be grounded in the distribution of the categories or scale points used by the observers.
3. It should constitute a numerical scale between at least two points with sensible reliability interpretations.
4. It should be appropriate to the level of measurement of the data.
5. Its sampling behaviour should be known or at least computable.

In scientific literature, the tasks in which unreliable data must be filtered are usually solved by statistical methods. Data whose deviation from the mean exceeds certain values are filtered. This paper proposes an improved method for filtering unreliable data; two types of preferences applied to a type of raw data. In this case, the statistical methods are not applied to the initial data but to the value of the distance function between the two preferences. This allows for the disclosure of conflicting or erroneous data.

Errors in the survey data may have resulted from both simple negligence and a completely irresponsible approach to the survey. Another possible reason for the erroneous data is that the respondents did not understand the rules for completing the questionnaire.

The study also presents the calculation of group preferences. All prioritisation methods are designed to "average" the opinions of all members of the group and to obtain a collective opinion. The peculiarity of the proposed method is that it is adapted to the research methodology. Participants' opinions were averaged, and two types of preferences were obtained. The same types of preferences were used when collecting the data. A comparison of the extent to which the "collective opinion" obtained by the proposed method is close to that of the participants in the group requires further research. This would be the subject of another study requiring comparison with other methods, such as Kemeny median, voting methods, and similar methods of constructing collective opinion.

6. Conclusions

This study presents a case of evaluating and ranking 11 values (life goals) using a questionnaire filled in by Lithuanian and Chinese students (male and female). The main goals of the research were to propose the methodology and tool for reliable data selection, apply the filter to the data and group preferences of Lithuanian and Chinese students (separately for women and men), and compare the corresponding group preferences. Four data sets were used for the research, two Lithuanian and two Chinese data sets.

It was shown that the data sets of the students from China filtered at least twice as much data as the Lithuanian students' data sets: 20–24% were left in Chinese men's data, 25–33% were in Chinese women's data; in comparison, 51.9–57% were left in Lithuanian men's data, 58.8–60% were left in Lithuanian women's data. Lithuanian students' data were more reliable than Chinese students' data. Another observation was that women's data were more reliable than men's (both for Lithuanian and Chinese students). It shows that men have greater variability in assessments than women, a finding that requires more in-depth research.

After applying the proposed methodology, the ranking and evaluation methods yielded similar results. All students (Lithuanian and Chinese, male and female) rated the most important values similarly. Highly ranked values in all data sets were: x5-family, x6-

love, x3-health, and x1-business (more important for men). The data set x4-social/friends was relatively highly rated by Lithuanian men. The data sets x8-contribution, x10-spiritual, x11-self-image (less important for Lithuanian and Chinese men), and x7-recreation/fun were rated the lowest. The data set x2-finance was less important to Chinese students than Lithuanian students. In conclusion, the data of Lithuanian and Chinese respondents are reliably similar, with minimum deviations detected in the ranking results.

Author Contributions: Conceptualization, A.K. and S.D.; methodology, A.K. and S.D.; software, A.K.; validation, A.K., N.K. and S.D.; formal analysis, N.K.; investigation, N.K. and S.D.; resources, S.D.; data curation, A.K.; writing—original draft preparation, S.D.; writing—review and editing, N.K. and S.D.; visualization, N.K.; supervision, A.K.; project administration, S.D.; funding acquisition, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted by the guidelines of the Declaration of Helsinki. Ethical review and approval were waived for this study, due to the methods used in this study (research type-sociological, anonymous online questionnaire, no direct contact with participants of this study, voluntary participation).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

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

References

1. Wedel, M.; Kannan, P.K. Marketing analytics for data-rich environments. *J. Mark.* **2016**, *80*, 97–121. [CrossRef]
2. Jaskó, S.; Skrop, A.; Holczinger, T.; Chován, T.; Abonyi, J. Development of manufacturing execution systems in accordance with Industry 4.0 requirements: A review of standard-and ontology-based methodologies and tools. *Comput. Ind.* **2020**, *123*, 103300. [CrossRef]
3. Verhoef, P.C.; Broekhuizen, T.; Bart, Y.; Bhattacharya, A.; Dong, J.Q.; Fabian, N.; Haenlein, M. Digital transformation: A multidisciplinary reflection and research agenda. *J. Bus. Res.* **2021**, *122*, 889–901. [CrossRef]
4. Nambisan, S.; Lyytinen, K.; Majchrzak, A.; Song, M. Digital Innovation Management: Reinventing innovation management research in a digital world. *Manag. Inf. Syst. Q.* **2017**, *41*, 223–238. [CrossRef]
5. Bleier, A.; Goldfarb, A.; Tucker, C. Consumer privacy and the future of data-based innovation and marketing. *Int. J. Res. Mark.* **2020**, *37*, 466–480. [CrossRef]
6. Arena, F.; Pau, G. An overview of big data analysis. *Bull. Electr. Eng. Inform.* **2020**, *9*, 1646–1653. [CrossRef]
7. Pennathur, P.R. A theoretical framework for understanding creator-consumer information interaction behaviors in healthcare documentation systems. *Appl. Ergon.* **2020**, *84*, 103034. [CrossRef] [PubMed]
8. Reason, J. *Human Error*; Cambridge University Press: London, UK, 1990.
9. Zhang, D. Data Reliability Assessment Based on Subjective Opinions. Ph.D. Thesis, University of Pittsburgh, Pittsburgh, PA, USA, 2021.
10. Wang, Y.; Wang, J.; Yao, T.; Li, M.; Wang, X. How does social support promote consumers' engagement in the social commerce community? The mediating effect of consumer involvement. *Inf. Process. Manag.* **2020**, *57*, 102272. [CrossRef]
11. Martin-Delgado, J.; Martínez-García, A.; Aranaz, J.M.; Valencia-Martín, J.L.; Mira, J.J. How much of root cause analysis translates into improved patient safety: A systematic review. *Med Princ. Pract.* **2020**, *29*, 524–531. [CrossRef]
12. Yadav, N.; Shankar, R.; Singh, S.P. Cognitive aspects of lean six sigma. *Qual. Quant.* **2021**, 1–60. [CrossRef]
13. Margaris, D.; Vassilakis, C.; Spiliotopoulos, D. What makes a review a reliable rating in recommender systems? *Inf. Process. Manag.* **2020**, *57*, 102304. [CrossRef]
14. Dalvi-Esfahani, M.; Alaedini, Z.; Nilashi, M.; Samad, S.; Asadi, S.; Mohammadi, M. Students' green information technology behavior: Beliefs and personality traits. *J. Clean. Prod.* **2020**, *257*, 120406. [CrossRef]
15. Beerens, R.J.J.; Tehler, H.; Pelzer, B. How can we make disaster management evaluations more useful? An empirical study of Dutch exercise evaluations. *IJDRR* **2020**, *11*, 578–591. [CrossRef]
16. Galvin, D.J.; Seawright, J.N. Surprising causes: Propensity-adjusted treatment scores for multimethod case selection. *Sociol. Methods Res.* **2021**, 00491241211004632. [CrossRef]
17. Schanze, J.L. Response behavior and quality of survey data: Comparing elderly respondents in institutions and private households. *Sociol. Methods Res.* **2021**, 0049124121995534. [CrossRef]
18. Whitley, B.E., Jr.; Kite, M.E. *Principles of Research in Behavioral Science*; Routledge: New York, NY, USA, 2012.

19. Taherdoost, H. Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. How to test the validation of a questionnaire/survey in a research. *SSRN* **2016**. [CrossRef]
20. Krylovas, A.; Dadelienė, R.; Kosareva, N.; Dadelo, S. Comparative evaluation and ranking of the European countries based on the interdependence between human development and internal security indicators. *Mathematics* **2019**, *7*, 293. [CrossRef]
21. Krylovas, A.; Kosareva, N.; Dadelo, S. European countries ranking and clustering solution by children's physical activity and human development index using entropy-based methods. *Mathematics* **2020**, *8*, 1705. [CrossRef]
22. Krylovas, A. *Diskrečioji Matematika. [Discrete Mathematics]*; Technika: Vilnius, Lithuania, 2009. (In Lithuanian)
23. Atherton, O.E.; Grijalva, E.; Roberts, B.W.; Robins, R.W. Stability and change in personality traits and major life goals from college to midlife. *Pers. Soc. Psychol. Bull.* **2021**, *47*, 841–858. [CrossRef]
24. Roberts, B.W.; O'Donnell, M.; Robins, R.W. Goal and personality trait development in emerging adulthood. *J. Personal. Soc. Psychol.* **2004**, *87*, 541. [CrossRef]
25. Chua, C. *Personal Excellence Book 2*; Personal Excellence: Singapore, 2012.
26. Dadelo, S.; Turskis, Z.; Zavadskas, E.K.; Kačerauskas, T.; Dadelienė, R. Is the evaluation of the students' values possible? An integrated approach to determining the weights of students' personal goals using multiple-criteria methods. *Eurasia J. Math. Sci. Technol. Educ.* **2016**, *12*, 2771–2781. [CrossRef]
27. Dadelo, S. Students' personal life goals assessed using settings weighting method. In *Managing Intellectual Capital and Innovation for Sustainable and Inclusive Society: Managing Intellectual Capital and Innovation, Proceedings of the MakeLearn and TIIM Joint International Conference, Bari, Italy, 27–29 May 2015*; ToKnowPress: Lublin, Poland, 2015; pp. 1941–1945.
28. Parsons, S.; Kruijt, A.W.; Fox, E. Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Adv. Methods Pract. Psychol. Sci.* **2019**, *2*, 378–395. [CrossRef]
29. Lilienfeld, S.O.; Strother, A.N. Psychological measurement and the replication crisis: Four sacred cows. *Can. Psychol./Psychol. Can.* **2020**, *61*, 281–288. [CrossRef]
30. Hayes, A.F.; Krippendorff, K. Answering the call for a standard reliability measure for coding data. *Commun. Methods Meas.* **2007**, *1*, 77–89. [CrossRef]