

Automated System for Determining a Damage Class for Sections of a Wastewater Disposal Network

Olena Shapovalova^a, Olga Starkova^a and Ganna Solodovnyk^a

^aKharkiv National University of Civil Engineering and Architecture, Sumska street, 40, Kharkiv, 61002, Ukraine

Abstract

The purpose of the study is to provide rationale and describe the process of designing an automated system for determining a damage class for sections of a wastewater disposal network using one of the up-to-date algorithms of data mining. The current stage of development of society requires the introduction of mathematical methods and mining models and the latest developments in the field of IT technology, even in such a well-established industry as public utilities. The proposed automated system provides the user with the opportunity to distribute sections of the network by clusters with the determination of their centers with the subsequent prioritization of repair work. The paper proposes a new approach to determining a damage class for sections of a wastewater disposal network using clustering algorithms with the option of choosing one of two metrics (Manhattan distance and Euclidean metric). The software component implemented in the Java programming language with the use and within the operation of the system allows determining by a number of section's criteria their affiliation with a certain class with the option of further ranking and prioritization of renovation work. The software is a web application which, on condition of good Internet connection, can be useful for professionals engaged in maintaining trouble-free operation of wastewater disposal networks at any time and at any location.

Keywords 1

Damage class, wastewater disposal network sections, clustering, web application

1. Introduction

Among the challenges faced by public utilities in cities and localities across Ukrainian, it is the issues of ensuring the smooth operation of wastewater disposal networks that head the list [1-3]. Sewer canals, the surfaces of which are constantly exposed to aggressive substances, over time suffer from typical damage, which results in the failure of certain sections and the occurrence of an emergency situation. A timely diagnostics of the most vulnerable sections and the network as a whole, followed by preventive measures to maintain the reliable operation of the system may significantly reduce the cost of repair in the event of an accident and ensure the provision of quality services to consumers [4-6].

Automating the processing of survey data about the condition of canals and ranking sections by the nature and extent of their damage is a relevant task at the moment, which will minimize the associated labor costs. Involving the latest developments in mathematical modeling, data mining and IT technology in solving practical problems, in particular in the construction, renovation and maintenance of complex systems [7-8] is a common practice among modern researchers trying to find optimum solutions and reduce costs through implementation of high technology [8-10].

The authors of a number of papers [5-10] suggest different approaches to deal with the issue. For instance, the paper [5] considers the issues of the strategy of selection of potential objects for rehabilitation of water supply networks based on the analysis of the results of technical diagnostics of the condition of the sections according to a number of appropriate criteria. In this case, the criteria are the dynamics of changes in the reliability of particular sections, the presence of external destabilizing

ITTAP'2021: 1nd International Workshop on Information Technologies: Theoretical and Applied Problems, November 16–18, 2021, Ternopil, Ukraine

EMAIL: shapovalova.olena@kstuca.kharkov.ua (A. 1); starkova@kstuca.kharkov.ua (A. 2); solodovnik@kn-it.info (A. 3)

ORCID: 0000-0003-4566-6634 (A. 1); 0000-0002-9034-8830 (A. 2); 0000-0001-6323-5083 (A. 3)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

factors, age, technical condition and repairability of pipelines, likely residual useful life period, preliminary costs of rehabilitation, the ratio of the actual cost of pipelines to the required costs in case of their rehabilitation, restrictions on financial costs.

The method and information system proposed in the paper [6] for determining the classification of the sewer network sections by condition allows determining the condition class of a sewer section, rating points, and lists of both individual sections and entire sewers that need priority rehabilitation. Unfortunately though, the effective use of this method requires special training and a high level of mastery of special software tools, which often does not allow fully applying it in the practice of planning repair and rehabilitation work.

The information and technical support offered by the authors allows storing the results of the technical diagnostics of the condition of sections in the data bank and, retrieving, if necessary, the necessary data on request to evaluate the listed criteria and determine the area of the urban water supply network or the area with the highest accident rate in pipelines (according to the diameters, materials and service life selected for analysis) [5, 6].

Attempts made by domestic researchers to assess the condition and elaborate a priority list for rehabilitation of wastewater disposal network facilities have been quite successful [7-10]. The information system proposed by the authors allows automating the process of elaborating a list of priority facilities of a wastewater disposal network to be rehabilitated, and visualizing, based on the data about the condition of the sewer network, the network section which the user is interested in with all its appropriate parameters [9].

2. Materials and methods of research

The reviewed studies do not give the unambiguous and general description which would allow determining the classification of the sewer network sections, being guided by their technical condition. Therefore, it is of relevance to consider the existing methods for determining the classification of sewers by condition. Moreover, at the moment there are no automated programs for determining the damage class using cluster analysis [11-13].

The relevance of the study and the development of a software application based on the study's findings are confirmed by the permanent interest of researchers in the issue [1-10] and involve the need for creating an automated system for determining the damage class of the wastewater disposal network sections. This application will be useful to specialists in the area of maintenance of trouble-free operation of the wastewater disposal network and will provide an opportunity to best develop repair plans based on the analysis of large amounts of information and classification of sections according to the level of their damage.

The study aims at providing rationale and describing the process of designing an automated system for determining a damage class for sections of a wastewater disposal network using a clustering algorithm [14-16]. The automated system is meant to distribute the wastewater disposal network sections by clusters (with the determination of their centers) with the subsequent prioritization of repair work [16-17]. The object of research is a software application for the implementation of clustering algorithms to differentiate sections of the wastewater disposal network according to the extent of their damage. The subject of research is the technology for implementing algorithms of analysis for the wastewater disposal network sections with the use of the IntelliJ Idea integrated development environment tools and Java programming language.

3. Research findings

According to the results of the analysis of the subject area, the input data for the automated system for determining the damage class of a wastewater disposal network section included the name of the section and parameters of cross section reduction, cracks, deformation, pipe ruptures, reinforced concrete corrosion, and the output data included the affiliation of the section with a particular cluster according to the appropriate parameters of its center. The user is given the opportunity to initiate, according to the input data, the calculation process, review and download the data to a file. In the course of loading the input data, the input information is prepared in a certain way and loaded into the system for further analysis and identification of consistent patterns.

The initiated calculation process activates the data clustering process, during which the user selects the metrics, the number of clusters and iterations and receives the distribution of sections by clusters [18-20]. The calculation results in the form of a file with information about each section with the indication of the damage class, and the coordinates of the centroids of clusters are stored and can be reviewed at the user's request and used in the future to make decisions on prioritization of repair work[20-23].

An activity diagram shows the sequence of actions of the automated system and the division of sections into clusters (Fig. 1).

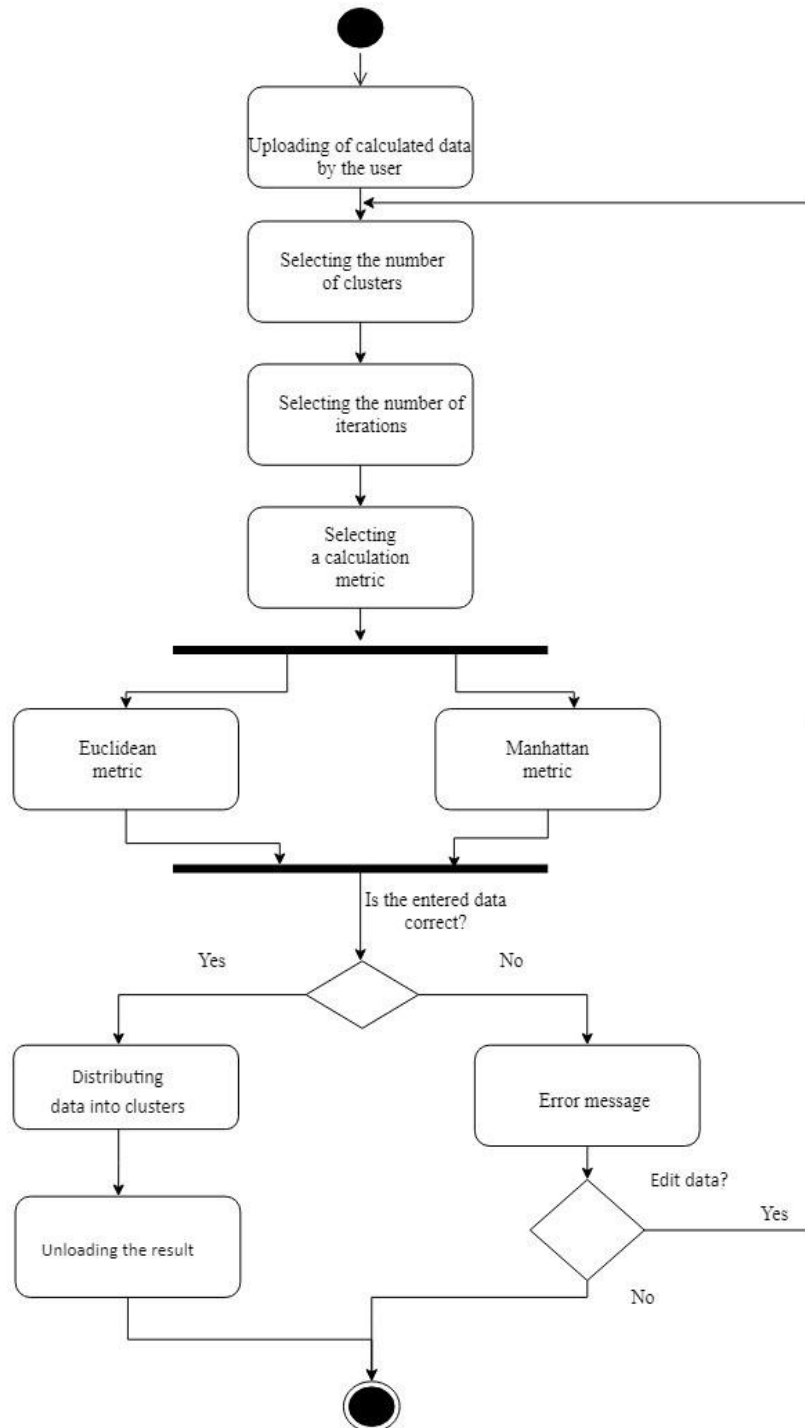


Figure 1: Activity diagram

In the course of operation, the system performs the algorithmic and mathematical calculations as follows:

- Forming a file from the output data;
- Cluster distribution of sections selected for review using the k-means clustering method [24-25];
- Searching for cluster centroids using Euclidean metric [26] and Manhattan metric [27].

To perform data clustering using the k-means clustering method, the number of clusters and the number of iterations are set automatically. Each object is represented by a sequence of standardized attributes, i.e. the values of all parameters (attributes) are reduced to the interval [0,100]. In the course of operation the algorithm of the object is given the cluster to which centroid it appears to be the nearest. The described operations are performed until the number of operations exceeds the maximum allowable number. An additional prerequisite to exiting the loop is to reach a state when objects are no longer assigned to new clusters. At the output of the algorithm, the user is provided with a set of objects divided into clusters, the centroids of which are equidistant from each other.

The diagram of the classes is shown in Fig. 2

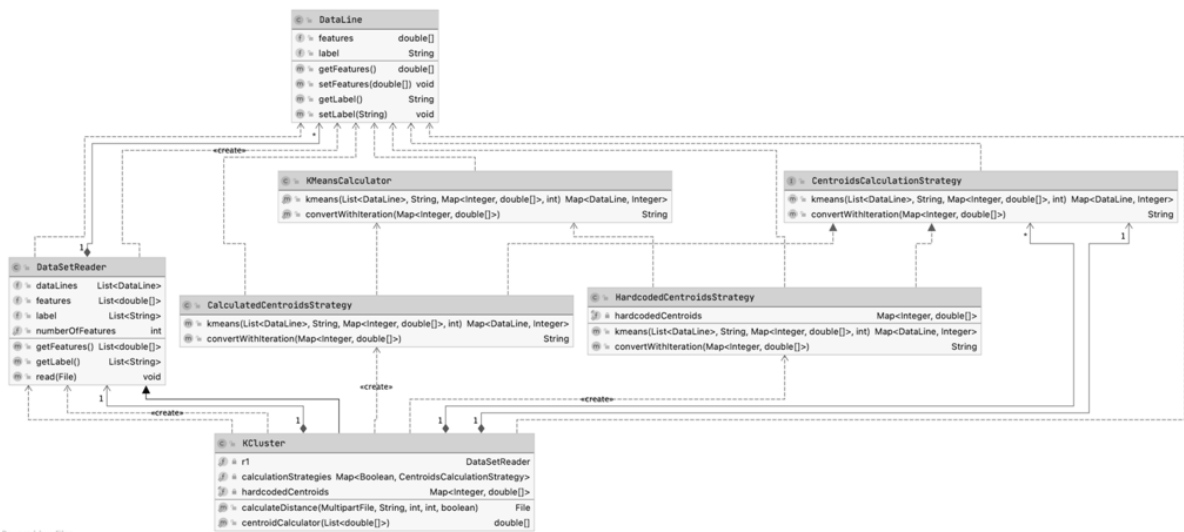


Figure 2: The diagram of the classes, part 1

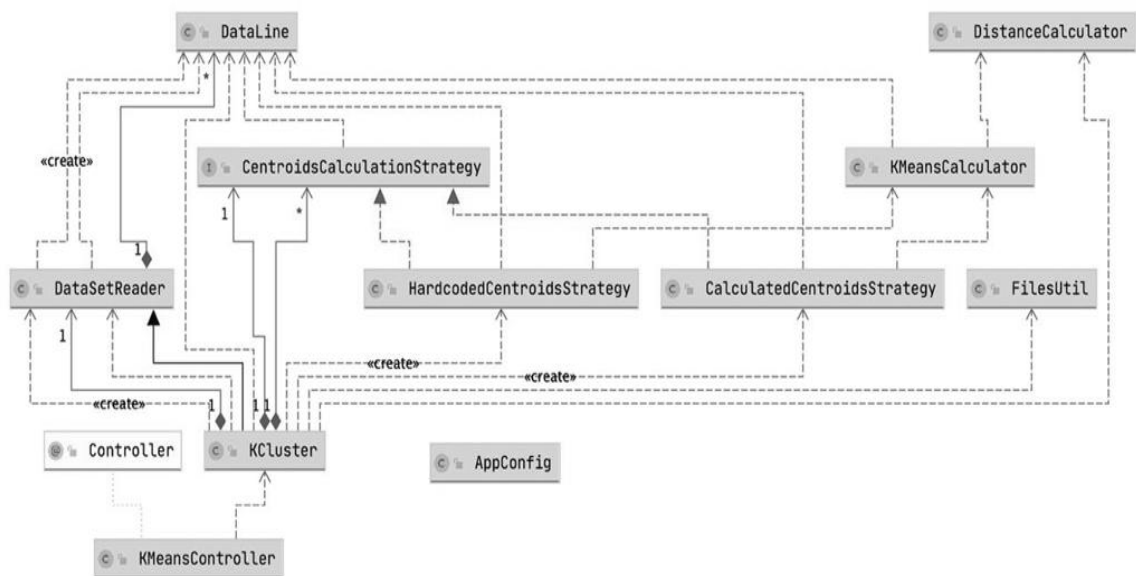


Figure3: The diagram of the classes, part 2

The classes of the web application are placed in 5 packages, which are responsible for calculations, configuration, display and application layers, such as Model, Service and Controller (Fig. 4).

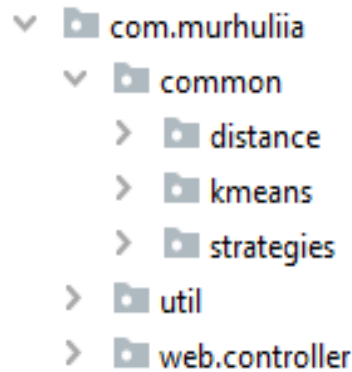



Figure4: Batch structure of the application

The following tools were used to create the web application:


- Apache Maven framework, designed to automate the collection of projects based on the description of their structure in files in the POM (Project Object Model) language;.
- Tomcat open source servlet container, which allows running web applications and contains a number of self-configuring programs;
- Java TestNG programming language testing system;
- Library for modular testing of software in Java JUnit and its extension DBUnit;
- A set of Bootstrap tools for creating websites and web applications;
- Spring Boot Java-based open source environment.

To start working with the web application, the user should go to the main page of the website (Fig.5).


Determining a damage class for sections of wastewater a disposal network



K-means clustering method



Euclidean metric



Manhattan metric

Clustering algorithm

K-means

Select data

File is not selected

Euclidean
 Manhattan

Number of clusters Number of iterations Centroid calculation

Figure 5: The window of the main page of the website

In this form, the user can choose the number of clusters, the number of iterations, calculate or not the centroids, apply the Euclidean or Manhattan metrics. Also, a .txt file with calculation data must be uploaded.

The uploaded file must contain in the lines the numerical data of the section in terms of 5 parameters (cross section reduction, crack, deformation, pipe rupture, and reinforced concrete corrosion) and its text name, separated by commas. If any parameter is missing, it is acceptable to replace it with the number 0.

Clicking the “Calculate” button after uploading the input data will get the result. For ease of further use and analysis, the data can be presented in tabular form using the Excel program built-in tools (Fig. 6).

4. Discussion of the findings

The developed automated system is used to determine the damage class for sections of a wastewater disposal network. The system is based on the main criteria, principles and methodological tools for classifying potential objects for renovation, which are covered in the research papers [2,3] and cluster analysis algorithms, the area of application of which is wide enough and discussed in the papers [11-14].

	A	B	C	D	E	F	G	H	I
	Cross section reduction,%	Crack, mm	Deformation, %	Pipe rupture, sm	Corrosion, %	Section,	Cluster,	number	
2	5,2	0,99	9,5	0	28,1	37	3		
3	39,2	15,2	49,9	0	92,2	7	0		
4	8,2	1,1	7,6	0	26,2	33	3		
5	36	16	46	0	96	2	0		
6	11,1	4,9	12,3	0	49,1	23	2		
7	3,3	0,11	6,6	0	7,4	42	4		
8	16	35	15	0	40	21	2		
9	7,7	1,21	6,4	0	19,2	38	3		
10	21,9	7,4	33,2	2,1	66,6	18	1		
11	7,5	1,25	8	0	20	31	3		
12	5,2	14	9,1	0	23,5	32	3		
13	18,1	4	16,6	0	48	28	2		
14	4,2	0,37	4,9	0	8,2	46	4		
15	37,1	13,1	46,2	8,2	88,8	6	0		
16	8,9	0,89	8,2	0	25,2	36	3		
17	42,2	15,2	48,1	10	81,6	10	0		
18	19	4	16,2	0	42,1	26	2		
19	23,5	7,3	32,1	1,8	52,2	16	1		
20	8,1	1,11	7,2	0	22,1	39	3		
21	2,5	0,25	3	0	6	41	4		
22	1,10	0,33	2,5	0	8,2	43	4		
23	36,2	16,3	49,6	9,1	89	3	0		
24	1,6	0,27	3,2	0	6,9	47	4		
25	25	7,2	27,9	3,2	60	11	1		
26	4,4	0,09	5,2	0	2,2	48	4		
27	3,6	0,19	5,2	0	9,1	45	4		
28	1,7	0,33	5,1	0	8,2	49	4		
29	22,5	6,2	33,1	2,5	56,2	12	1		
30	28,1	7	35	2,5	65	20	1		
31	24,1	9,2	36,1	4,2	59,9	17	1		
32	13,3	3,2	15,2	0	47,2	27	2		
33	18,1	3,1	17,5	0	39,9	25	2		
34	16,2	2,6	14,4	0	33,3	24	2		
35	16,2	4,7	13,2	0	41,2	29	2		
36	42,5	19,2	47,7	7,7	87	8	0		
37	29,2	9,3	39,6	1,4	50,2	13	1		
38	28,1	8,2	34,2	2,5	62	15	1		
39	1,9	0,31	4,2	0	62	50	4		
40	32,2	18,1	52,4	7	99,1	4	0		
41	9,1	1,33	6,6	0	15,2	40	3		
42	1,8	0,42	1,2	0	4,2	44	4		
43	21,4	6,2	36,2	3,5	64	14	1		
44	41,7	16,6	47	8,1	80	9	0		

Cluster	Centroids
0	[30.0, 10.0, 40.0, 7.6, 85.0]
1	[25.0, 7.5, 30.0, 2.5, 60.0]
2	[16.0, 3.5, 15.0, 0, 40.0]
3	[7.5, 1.25, 0, 0, 20.0]
4	[2.6, 0.25, 3.0, 0, 5.0]

Figure 6: The distribution of sections by clusters in tabular form

1. Conclusions

According to the findings of the study, using the example of well-known automated systems for determining the damage class for the wastewater disposal network sections, approaches to dealing with the issue were considered, the subject environment was analyzed and an activity diagram was plotted, which was the basis for designing an automated system. Based on the analysis of the subject area, the input and output data for the system were selected; a cluster analysis was chosen as the method for determining the priority of repair work.

A system has been designed and implemented, the software component of which is a web-application with client and server parts that automate the determination of the damage class for the wastewater disposal network sections.

The user is given the opportunity to store the results of calculations in the database and perform their further processing using spreadsheets. The system was tested and showed high speed of statistical data processing and 100% accuracy of distribution of sections by clusters on the test sample. In the future, it is planned to use the designed system for the analysis of sections of a wastewater disposal network both in real time and for drawing up plans for renovation work.

2. References

- [1] Honcharenko D.F., Bondarenko A.I., Bulgakov V.V. and Garmash A.A. On the issue of ensuring the maintainability of sewer tunnels in Kharkov. *Naukovyi visnyk budivnytstva*, 2016. Issue 2, pp. 144–148.
- [2] Starkova O.V. Models of a reasonable choice of the method of repair and restoration of a section of the sewer network. *Naukovyi visnyk budivnytstva*. 2016. Issue 3, pp. 80-85.
- [3] Starkova O.V., Shapovalova Ye.A., Gnuchikh L.A. Modeling the choice of the method of restoration of sewerage networks. *Komunalne hospodarstvo mist*. 2008. Issue 85, pp. 19 –26.
- [4] Bulgakov Yu.V. Research of the process of destruction of the construction of the sewer tunnel collector. *Naukovyi visnyk budivnytstva*, 2015. Issue 5, pp. 79-84.
- [5] Orlov V.A., Kharkin V.A. Development of a strategy for the restoration of wastewater disposal networks. *Stroitelstvo i Arkhitektura*, 2001, pp. 15-32.
- [6] Korynko Y.V., Starkova O.V., and Shevchenko A.A. Methodological foundations of computer modeling of wastewater disposal systems. *Naukovyi visnyk budivnytstva*, 2003. Issue 23, pp. 223–229.
- [7] Honcharenko D.F., Starkova O.V. and Aleynykova A.Y. Development of an automated system for selecting a method for restoring water pipelines using a fuzzy logic. *Systemy obrobky informatsiyi*, 2014. Issue 8 (124), pp. 18-23.
- [8] Solodovnyk H.V., Deyneha A.A. Making multi-stage decisions using information technology. *Naukovyi visnyk budivnytstva*, 2019. Issue 2, pp. 165-166.
- [9] Starkova O.V., Shapovalova E.A., Hnuchykh L.A. and Bondarenko D.A. Development of an automated information system for determining priority objects for the renovation of water supply pipelines. *Komunalne hospodarstvo mist*, 2011. Issue 99, pp. 312-316.
- [10] Honcharenko D., Shumakov I., Starkova O., Aleinikova A. and Mikautadze R.. Methodological and computer-based support for choosing underground utility networks renovation method // *MATEC Web of Conferences* 230, 02010, 2018.
- [11] Quackenbush J. Computational analysis of microarray data. *Nature reviews genetics*, 2001, vol. 2, no. 6, pp. 418–427.
- [12] Sugar C. A., James G. M. Finding the number of clusters in a dataset. *Journal of the American Statistical Association*, 2003, vol. 98, no. 463, pp. 750–763.
- [13] Tibshirani R., Walther G. Cluster validation by prediction strength. *Journal of Computational and Graphical Statistics*, 2005, vol. 14, no. 3, pp. 511–528.

- [14] Tibshirani R., Walther G., Hastie T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 2001, vol. 63, no. 2, pp. 411–423.
- [15] Cuevas A., Febrero M., Fraiman R. Cluster analysis: a further approach based on density estimation. *Computational Statistics & Data Analysis*, 2001, vol. 36, no. 4, pp. 441–459.
- [16] Stuetzle W. Estimating the cluster tree of a density by analyzing the minimal spanning tree of a sample. *Journal of classification*, 2003, vol. 20, no. 1, pp. 25–47.
- [17] Pelleg D., Moore A. W. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *ICML*, 2000, pp. 727–734.
- [18] Volkovich Z., Brazly Z., Toledano-Kitai D., Avros R. The Hotelling’s metric as a cluster stability measure. *Computer modelling and new technologies*, 2010, vol. 14, pp. 65–72.
- [19] Barzily Z., Volkovich Z., Akteke-Ozturk B. On a minimal spanning tree approach in the cluster validation problem. *Informatika, Lith. Acad. Sci.*, 2009, vol. 20, no. 2, pp. 187–202.
- [20] Hamerly Y. F. G. PG-means: learning the number of clusters in data. *Advances in neural information processing systems*, 2007, vol. 19, pp. 393–400.
- [21] Dudoit S., Fridlyand J. A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome biology*, 2002, vol. 3, no. 7, research0036. Available at: <http://www.genomebiology.com/> (assessed: 17.08.2021).
- [22] Lange T., Roth V., Braun M. L., Buhmann J. M. Stability-based validation of clustering solutions. *Neural computation*, 2004, vol. 16, no. 6, pp. 1299–1323.
- [23] Ben-Hur A., Elisseeff A., Guyon I. A stability based method for discovering structure in clustered data. *Pacific symposium on biocomputing*, 2002, vol. 7, no. 6, pp. 6–17.
- [24] Adam Coates and Andrew Y. Ng. *Learning Feature Representations with K-means*, Stanford University, 2012.
- [25] Mirkes E.M. *K-means and K-medoids applet*. University of Leicester, 2011.
- [26] Dyuran B., Odell P. *Cluster analyze*. Moscow: Statistika, 1977, 128 p.
- [27] Kim, J. O., Mueller, C. W., & Klecka, W. R. *Factorial, discriminant and cluster analyzes*. Moscow: Finansy i statistika, 1989, 215 p.