UDC 37.091.64:371.26:519.23

Olha Pronina
Candidate of Technical Sciences, Associate Professor of the Department of Computer Science
SHEI "Pryazovskyi State Technical University", Dnipro, Ukraine
ORCID ID 0000-0001-7085-8027
pronina_o_i@pstu.edu

Olena Brodniuk
Master of Computer Science
SHEI "Pryazovskyi State Technical University", Dnipro, Ukraine
lbrodniuk@ukr.net

Olena Piatykop
Candidate of Technical Sciences, Associate Professor of the Department of Computer Science
SHEI "Pryazovskyi State Technical University", Dnipro, Ukraine
ORCID ID 0000-0002-7731-3051
piatykop_o_ye@pstu.edu

Irina Fedosova
Doctor of Pedagogical Sciences, Professor of the Department of Computer Science
SHEI "Pryazovskyi State Technical University", Dnipro, Ukraine
ORCID ID 0000-0003-3923-8270
fedosova_i_v@pstu.edu

CLUSTER ANALYSIS OF EDUCATIONAL DATA FOR ASSESSING AND VISUALIZING EDUCATIONAL LOSSES IN SCHOOL EDUCATION: SOFTWARE IMPLEMENTATION

Abstract. The article is devoted to the urgent problem of educational losses that arose as a result of the COVID-19 pandemic and military operations in Ukraine. The article considers the concept of educational losses, the prerequisites and causes of their occurrence, as well as the categories of losses. The experience of other scientists in measuring and compensating for educational losses is also analyzed. Losses in learning and the decline in the level of students' academic achievements are analyzed separately. The authors propose a comprehensive approach to measuring and compensating for educational losses using cluster analysis methods. The k-means and hierarchical clustering algorithms were applied, which made it possible to identify groups of students with similar academic achievements. Python, with the Pandas, Matplotlib, and Scikit-learn libraries, was used for data processing and analysis. The graphical interface was implemented to visualize the results and facilitate teachers' work. The article presents experimental studies on grouping students 5-6 grades into clusters, as well as loss calculations obtained using the developed software. The results confirmed the existence of a stable correlation between achievements in different subjects and made it possible to calculate the level of individual educational losses.

The proposed approach creates conditions for personalization of learning, timely diagnosis of gaps and planning of compensatory measures, including differentiated tasks, group tutoring and additional psychological support. It was concluded that the automation of educational data analysis and the use of cluster analysis significantly increase the effectiveness of overcoming educational losses and contribute to ensuring equal access to quality education even in crisis conditions.

This will allow teachers to better organize compensation for educational losses, adapt curricula to the individual needs of students.

Keywords: educational losses; distance learning; cluster analysis; individualization of learning; k-means; hierarchical clustering.

1. INTRODUCTION

The problem statement. Education is a key factor in preserving the rights of all people. From the point of view of development, education has significantly expanded human

© Olha Pronina, Olena Brodniuk, Olena Piatykop, Irina Fedosova, 2026.

capabilities, which has become one of the prerequisites for the sustainable (balanced) development of society [1]-[4]. Education and science have become the most principal tools for management, decision-making, and the formulation of civilization. Sustainable development is the basis for human development and the creation of conditions for the realization of its capabilities. The COVID-19 pandemic has had a significant impact on the education system around the world [5],[6]. In addition, the organization of the educational process in Ukraine is significantly complicated by a full-scale war [7], [8]. These factors have forced changes to traditional forms of education [5], [9], [10]. All this leads to many problems related to learning, in particular the tendency for students to lose out on learning [11], [12]. This problem is being studied by scientists in different countries [11], [13], [14]. Because learning losses are a significant problem that affects the education and life of students in general and disrupts the sustainable development of the country [15]. It can be defined as the difference between the expected level of educational achievement and the actual level achieved by a student.

Any form of overcoming learning losses is a control of knowledge and analysis of students' responses to tasks of varying complexity. Educational losses are learning losses associated with students' cognitive skills, knowledge gaps, failure to achieve the expected learning outcomes defined by educational programs, etc.[16].

Educational losses result from a range of factors and may be individual or systemic in nature, differing in scope and scale. [16]. During the pandemic, schools were closed around the world, and learning switched to a distance format. Many countries began to use a huge number of digital tools and approaches, while a new problem emerged: inequality between students with different resources. During the war in Ukraine, children experienced additional educational losses due to various factors, including military clashes and occupation, air raids, departure abroad and internal displacement within Ukraine. Another contributing factor was power outages caused by the massive shelling of critical infrastructure.

According to a study by the Ukrainian Center for Education Quality Assessment, educational losses may lead students to lose essential skills needed in adult life, reduce their potential for higher income, and ultimately affect society as a whole. [12]. But they also added that there are steps to cope with making up for educational losses. The following types of measures are being considered: teaching during the holidays; diagnostic work to measure learning losses during the summer period; adaptation of the educational program; development of additional tasks; methodological training of teachers; Summer Intensive online school.

Measuring and compensating for losses is a complex problem that requires the use of various research methods and the work of the teacher [17]. In the process of data analysis by the teacher, tasks of varying complexity are developed. And the teacher is not always able to process a large amount of information qualitatively [18]. Here, automation of the process of analyzing educational data using clustering is required. This can avoid various errors and reduce the complexity of the analysis, because it often happens that at one time or another the students` knowledge decreases.

It is very important to understand data on initial student knowledge, that is, information about what students knew and were able to do before the learning interruption. And to be able to compare this knowledge after a certain period of learning interruption. Compared to countries where educational losses are regularly measured, Ukraine faces significant challenges in this area. Before the pandemic and the war, the source of measuring success was the State final evaluation in grades 4 and 9, and in grade 11 the External independent evaluation. Since the beginning of the war, these types of control have simply not been carried out, except for the National Assessment of Physical Education, which greatly affects the objectivity of obtaining data on the success of education seekers. Therefore, various instruments – such as surveys, testing, and monitoring of educational quality – are used to assess different aspects of student learning. However, it is first necessary to diagnose students' levels of academic achievement

and learning success. Based on this diagnosis, strategies for overcoming educational losses can be developed. Given the range of indicators that influence learners, it is appropriate to examine the conditions contributing to learning losses and to develop a research methodology that incorporates cluster analysis for the examination of data on students' educational outcomes.

Analysis of recent studies and publications.

Since children are the future of any society, much attention is paid to their education, including many scientists who wonder about educational losses. So, Bilyakovska et al. [19] consider assessment as an improvement of the learning environment and adaptation to the needs of users in terms of cognitive, social and technical infrastructure. They also emphasize that the results of assessment may depend on various facts, such as social, cultural, and individual.

Szekely et al. [17] attempt to measure learning losses that have occurred during the pandemic. The study was conducted among children in low socioeconomic families in Mexico. The authors use a difference-in-differences approach and use similarity scores to find differences in student learning.

Aslam et al. [20] examine the development of universal e-learning in the context of artificial intelligence. The emphasis is on individualizing learning, which allows for better interaction with students and improved results. The study analyzes the evaluation of e-learning models using AI strategies, such as supervised and reinforced learning. The importance of various e-learning features for improving models is considered, as well as the effectiveness of algorithms, in particular deep learning methods, for large data sets.

Shesterova et al. [21] presents the overcoming of educational losses by performing sensory oriented tasks in physical education lessons. The authors consider one of the solutions to the loss of knowledge to be the use of physical exercises as a means of influencing the visual analyzers of education seekers. The work provides several examples of exercises aimed at this problem that directly affect the health of children. They can be used in any part of the lesson for a break or rest. The authors also provide recommendations in the work on the use of such exercises for primary school teachers and subject teachers. There are studies that have already explored the application of clustering to specific target groups in the field of education.

Matzavela and Alepis [22] rely on training data that can be classified using decision trees in the field of machine learning. The authors showed that the decision tree algorithm generates a classification model very well, which is easy to process and display graphically. The goal of the work is progress, that is, further improvement in student learning, which has the main goal of creating adaptive tests to assess the success of education seekers.

Gamboa Mora et al. [23] investigated the issue of implementing an appropriate model that allows predicting students at risk of losses, namely, in chemistry. The data was processed using Jupyter-Python, where a linear regression model was built based on tests (including the university entrance exam). It was concluded that linear regression is the best approach to measuring educational losses.

Tao et al. [24] present a study conducted by Chinese scholars that examines the clustering of educational resources based on data from 80 cities in China. The study reveals a trend toward increasing urban resilience in the central regions of the country, which the authors attribute to the concentration of educational resources in the urban areas of provincial capitals. The findings demonstrate that educational resource clustering has a significant impact on urban resilience, highlighting important implications for the optimal allocation of educational resources and the promotion of social justice in these cities.

Pronina and Piatykop [25] apply cluster analysis to predict students' academic performance during their studies at a higher education institution. At the first stage, the relationship between the results obtained during external independent testing and the subsequent distribution of scores across the relevant disciplines was checked. At the second

stage, a comparison was made between the predicted student grades and their actual grades obtained during their studies.

Research by Pamungkas et al. [26] aims to classify students of the Faculty of Computer and Information Education, Universitas Sebelas Maret, according to their learning styles, using the Felder and Silverman Model (FSLSM). The study, which included 58 students, used a quantitative descriptive method and cluster analysis. The results showed the formation of two main clusters with different characteristics of learning styles. The first cluster shows reflective, sensory, visual and global preferences, while the second has active, sensory, visual and global preferences. It assists teachers in choosing appropriate teaching strategies for each class.

Therefore, in order to investigate the compensation of educational losses, it is advisable to use cluster analysis to identify groups of students with similar educational losses and develop targeted strategies for compensating for learning losses.

The research goal. Analyzing the research, it was found that the optimal way to solve this problem is to use cluster analysis to measure and compensate for educational losses at the educational institution level. The research goal is increasing the efficiency of measuring and compensating for students' educational losses by applying cluster analysis algorithms and creating software tools to automate educational data processing, which will ensure the personalization of the educational process.

2. RESEARCH METHODS

Educational losses are combined not only with learning losses, but also with losses of opportunities in the comprehensive development of the child: this is the intellectual, psychological, emotional and social component.

Among the main approaches that are distinguished for the analysis of educational losses, there is an individual one, which involves the creation of an individual curriculum for a student with knowledge gaps. There is also a group approach, which aims to carry out compensatory measures for the entire class or a certain category of students.

Within the concept of educational losses, there are three components that are interconnected: learning losses (loss of skills and abilities); upbringing losses; losses in personal development [8,16]. To measure knowledge gaps, certain categories of educational losses are defined:

- learning losses loss of knowledge in individual subjects that have an impact on the students overall performance;
- psychological losses resulting from decreased motivation to study, which can form a negative attitude towards school;
- social losses that directly relate to social interaction in the classroom, school. They are reduced by limiting physical presence at school (distance learning);
- personal losses may include loss of self-confidence and decreased initiative.

Thus, measuring the level of learning losses is a priority task, without which it is impossible to work on overcoming learning losses in general.

The ways to compensate for educational losses are diagnostics of the student; analysis and assessment of primary knowledge; a work plan for compensation of knowledge; directly, compensation for educational losses; repeated diagnostics to make sure that the child has caught up with the necessary material.

To assess educational losses, a data sample was created that was generated directly at the level of the educational institution Nikolsk Secondary School of I-III Grades No. 1 named after A. D. Yakimenko of the Nikolsk Settlement Council of the Donetsk Region students of 5-6 grades. The sample included: students’ gender, race, parents’ education level, students’ diet,

information about taking a test preparation course, grades in mathematics, Ukrainian literature, and Ukrainian language. Mathematics, Ukrainian language (writing), and Ukrainian literature (reading) were selected for analysis as core components of the national curriculum. These subjects collectively assess analytical thinking, comprehension, and literacy skills, which are foundational across all disciplines. The distinction between reading (Ukrainian literature) and writing (Ukrainian language) was made to enable more precise identification of specific skill deficits.

While other subjects such as science or foreign languages are also crucial at the secondary level, this study focuses on core literacy and numeracy skills as the basis for identifying general learning losses. Future studies may expand the dataset to include additional disciplines. Before testing the model, pre-processing of the generated sample was carried out, namely: grouping student indicators into educational data groups, removing missing values and irrelevant records, converting percentages, scaling numerical data, and adapting categorical variables. At the beginning of the year, the sample looked like this, a fragment is presented in Table 1.

Table 1.

Fragment of the sample of students at the beginning of the2024 academic year

No	Gender	Race/Ethnicity	Test preparation course	Math score	Reading score	Writing score
1	female	groupD	completed	59	70	78
2	male	groupD	none	96	93	87
3	female	groupD	none	57	76	77
4	male	groupB	none	70	70	63
5	female	groupD	none	83	85	86
6	male	groupC	none	68	57	54
7	female	groupE	none	82	83	80
8	female	groupB	none	46	61	58
9	male	groupC	none	80	75	73
10	female	groupC	completed	57	69	77

After each topic, a test was conducted, the results of which were added to the sample for further calculation of learning losses. The calculation of learning losses is based on the results of educational measurements in individual subjects of students. For this purpose, a function has been created that checks the availability of data for analysis, the presence of at least two columns from each subject: math score (mathematics), reading score (Ukrainian literature), and writing score (Ukrainian language, writing), which are used for calculations.

This section describes the methodology, including the preparation and conduct of the research, as well as the experimental base of research, participants and investigating facilities. Also, the characteristics, criteria, indicators and parameters of assessment are provided. The names and registration numbers of scientific programs and research projects can be provided.

The calculation, itself learning losses, occurs according to a function that determines the maximum and minimum score in each subject by topic for each student. The percentage of the difference between the maximum and minimum score, relative to the maximum, is calculated using the following formula:

$$L_i = \frac{(a_{max} - a_{min})}{a_{max}} \cdot 100 \quad (1)$$

where L_i is learning losses for discipline i; a_{max} is the maximum score of the current subject; a_{min} is the minimum score of the current subject.

Resulting value learning losses for each student is added using the educational loss column. If the data are not found or loaded, the function will display an error message and, accordingly, the calculations will not occur.

The authors have developed software for clustering data to measure educational losses, allowing students to be grouped according to various indicators. For implementation, it was chosen to use the Anaconda development environment, the Python programming language, and a number of libraries. The Pandas library was chosen to work with data, the Matplotlib library was chosen to visualize data, and Scikit-learn was allowed to perform the cluster analysis itself directly. For interactive work with the user, an interface was created, the graphic design was created using Tkinter, which allows viewing the results in real time, shown in Figure 1.

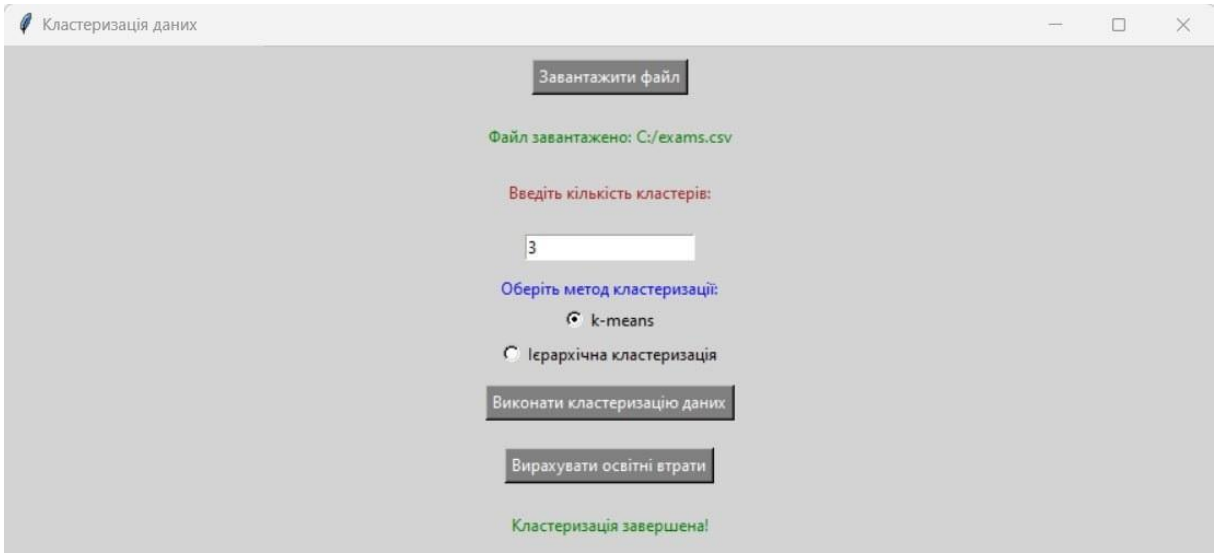

Figure 1. User interface of the developed software.

The first stage is the analysis and grouping of students according to key indicators – formed educational losses, which is performed automatically. The second stage, which is performed by the teacher directly based on the data received, introduces compensating measures to reduce educational losses. Thanks to grouping, it is easier for the teacher to adapt the material based on the similarity of the group of students. Grouping according to learning losses occurs using data clustering tools, in particular the k-means algorithm and hierarchical clustering.

The algorithms underlying the developed software product and grouping students by similarity are k-means and hierarchical clustering.

4. THE RESULTS AND DISCUSSION

The purpose of the experiment is to investigate the possibilities of using cluster analysis using Anaconda tools to measure learning losses of education seekers, using data analysis, K-means and hierarchical clustering algorithms.

In the first stage, after loading the data sample into the application, statistics are calculated for the sample for its comprehensive analysis.

Below are statistics for numeric columns for subject scores: number of items in the data set; mean for each score; standard deviation, which shows the spread of scores; minimum score value; percentile, that is, the value that the random variable sets but does not exceed it; 25% percentile, that is, about 59 points in all three tests for students; 50% is the median, which is

about 68–70 points; 75% percentile, which is about 79–80 points; the maximum value of the scores. The calculation for the current state of the sample is shown in Table 2.

Table 2.

Statistics calculated for the current state of the sample

	Math score	Reading score	Writing score
mean	67.810	70.382	69.140
std	15.250	14.107	15.026
min	15.000	25.000	15.000
25%	58.000	61.000	59.000
50%	68.000	70.500	70.000
75%	79.250	80.000	80.000
max	100.000	100.000	100.000

In the second stage, hierarchical clustering was constructed. Hierarchical clustering clearly demonstrates the class hierarchy and the optimal number of clusters for the current sample (Figure 2).

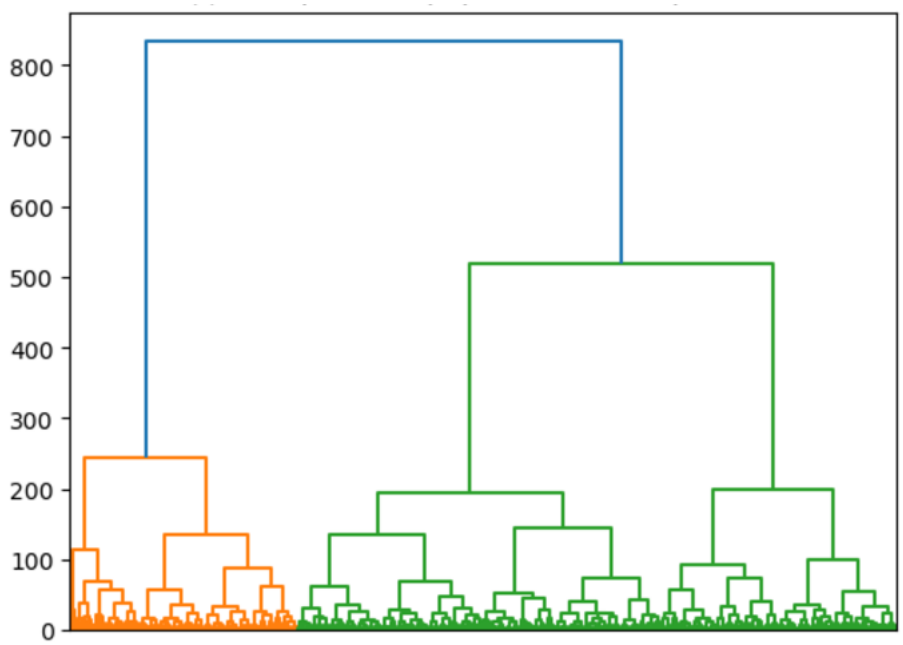

Figure 2. Results of hierarchical agglomeration clustering.

The vertical axis shows the similarity metric between objects, in this case the Euclidean distance. The horizontal axis shows each student in the sample as a separate cluster. As you move up the tree, students are grouped into clusters until all students are grouped into one large cluster.

The horizontal lines connecting the clusters indicate the distance between them. That is, we can say that the longer the line, the greater the distance between the merging clusters. And this means that the students in these clusters are less similar to each other. The graph itself shows how the data is structured in the form of a hierarchy, which allows you to visually select the optimal number of clusters based on the level of separation. In this case, you can see that three large clusters are optimally formed. This information is then used to build clustering using the K-Means method. Since the application provides the ability to enter the number of clusters

for K-Means clustering, this allows you to adjust the number of student groups. The results of clustering when using the K-Means algorithm are shown in Figure 3.

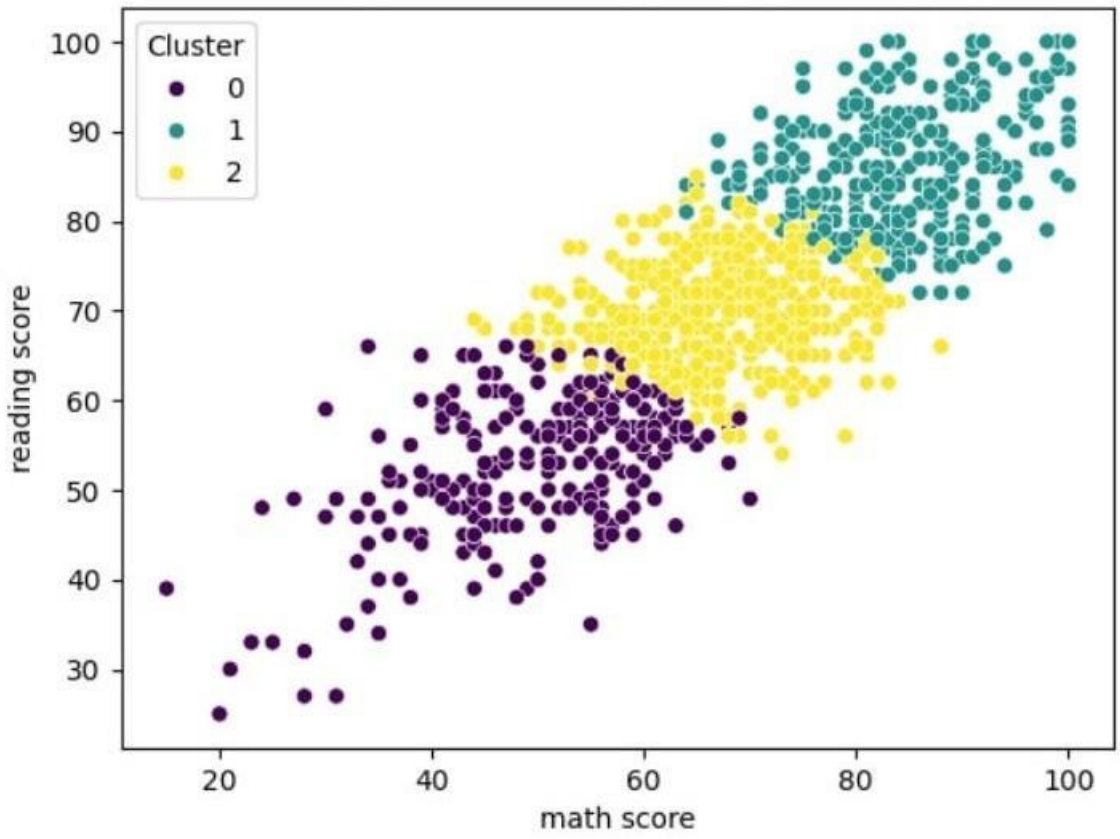

Figure 3. K-means clustering results.

Analyzing the obtained clustering results for students 5–6 grades, you can see that the distribution is made into 3 clusters, which are marked with colors (purple, green, yellow).

The graph shows the X-axis (math score) – math score and the Y-axis (reading score) – reading score. All points on the graph are students grouped by the similarity of their math and reading scores.

The clustering results showed three distinct groups of students with different levels of achievement. Group A (green color) showed consistently high scores in all subjects, Group B showed moderate scores with significant fluctuations in reading, and Group C (purple color) showed lower levels of achievement, indicating significant learning losses. These findings point to the need for differentiated instructional strategies.

That is, the clusters represent certain levels of academic achievement of students. From here, we see a trend –a positive correlation between the scores: students with high scores in mathematics have a high score in reading. This is also confirmed by the linear regression calculated and plotted in figure 4. The red line is the regression line that best describes the general trend of the relationship between them.

A study was also conducted to identify the relationships between scores in different subjects (student skills). Analyzing the results, it can be concluded that there is a pronounced correlation between three variables: Mathematics, Ukrainian language (writing) and Ukrainian literature (reading) scores. A sensitivity map was constructed, demonstrating that all coefficient values are in the range [0.79; 0.95], that is, there is a positive correlation between them. Students who have high scores perform equally well on all tasks.

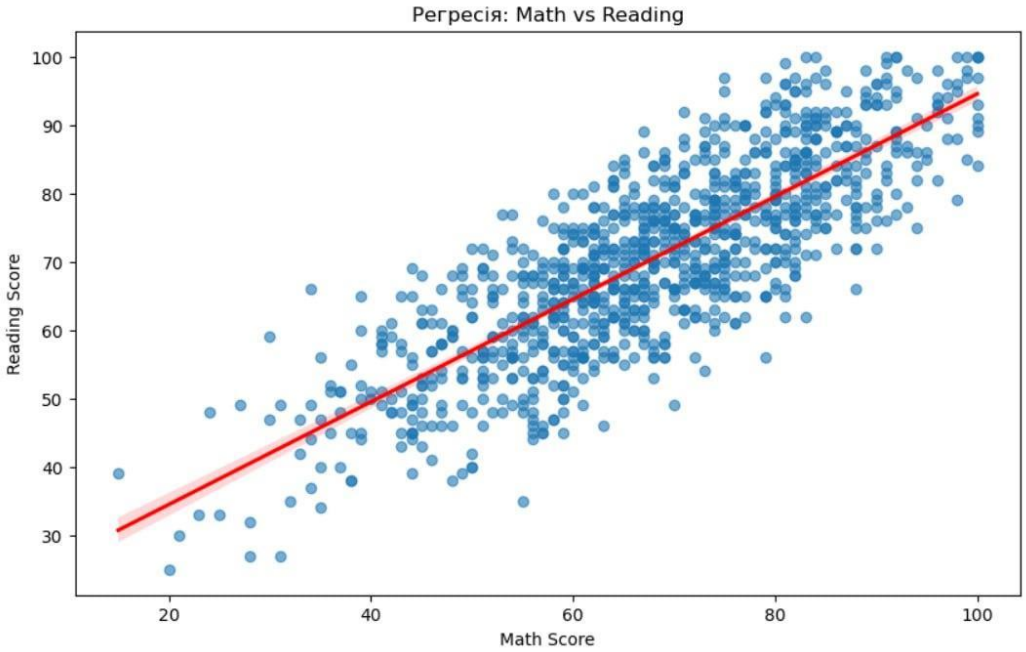

Figure 4. Sample regression line

At the last stage, the learning losses were calculated for each student for each subject. The total level of learning losses for each student was also calculated. A fragment of the sample with the calculation is shown in Table 3.

Table 3.

Sample fragment with calculation of total learning losses

No	Math score	Reading score	Writing score	Learning losses
1	59	2	3	4
5	96	57	70	83
70	93	76	70	85
78	87	77	63	86
24.36	9.38	25.97	10.00	3.49

The results of calculating learning losses for each student are a tool that allows the teacher to observe the dynamics of studying the discipline. This allows for timely response and the implementation of short-term or long-term activities to catch up on the material studied in parallel with the new material. Also, thanks to the calculation of educational losses, the teacher can predict the overall grade for the subject and timely inform parents about possible deterioration.

While learning gaps may naturally occur during the educational process, the scale and distribution of learning losses are significantly amplified by prolonged periods of distance learning and educational disruptions due to the COVID-19 pandemic and the ongoing military conflict in Ukraine. Using clustering with the developed software, it is possible to identify atypical concentrations of students with significant learning losses. This is consistent with documented national trends in educational disruptions during such crises.

Teachers can use the developed tool to quickly identify students who need more attention. The developed software allows teachers to prioritize students based on the severity of their learning loss and adapt the level of assignments accordingly.

Clustering will allow teachers to personalize learning. For example, students in different clusters could receive differentiated homework, feedback, or remedial sessions based on their specific patterns of learning loss.

Recommendations for addressing learning losses:

- implement targeted group tutoring based on identified clusters;
- provide additional resources and individual support for students in the cluster with the highest losses;
- conduct regular diagnostics using the developed software to monitor progress;
- train teachers to adapt instruction for heterogeneous student groups;
- integrate socio-emotional support for students showing motivational and psychological learning losses.

4. CONCLUSIONS AND PROSPECTS FOR FURTHER RESEARCH

The study revealed that education is a crucial factor in the sustainable development of society. Educational losses caused by the COVID-19 pandemic and military operations in Ukraine have serious consequences for the future of the younger generation.

The difficulties in the educational process caused by the crisis are due not only to the technical aspects of the transition to distance learning, but also to deeper social problems, such as inequality in access to quality education.

The developed software, which is based on cluster analysis methods, particularly k-means and hierarchical clustering algorithms, offers new opportunities to identify and better understand students' learning needs. The pilot study has only been conducted in one educational institution so far. Therefore, the results obtained primarily reflect the specific context of the school being analyzed. However, the proposed approach will allow for more effective personalization of the educational process, which is crucial for diagnosing and compensating for learning losses. The automation of educational data analysis contributes to reducing teachers' workload while improving the quality and adaptability of teaching.

Overall, to ensure equal access to quality education for all students, especially in times of crisis, urgent action is needed at all levels, from government policy to individual classroom approaches. Only through joint efforts can we overcome the challenges and ensure a successful future for all participants in the educational process.

REFERENCES (TRANSLATED AND TRANSLITERATED)

[1] O. Lavrentieva, V. Pererva, O. Krupskyi, I. Britchenko, and S. Shabanov, "Issues of shaping the students professional and terminological competence in science area of expertise in the sustainable development era," *E3S Web of Conferences*, vol. 166, p. 10031, 2020, doi: https://doi.org/10.1051/e3sconf/202016610031. (in English)

[2] E. Komarova and T. Starova, "Majority values of school biological education in the context of education for sustainable development," *E3S Web of Conferences*, vol. 166, p. 10029, 2020, doi: https://doi.org/10.1051/e3sconf/202016610029. (in English)

[3] K. P. Osadcha and N. V. Shumeiko, "Artificial intelligence, the labor market, and education for sustainable development: the points of intersection," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1415, no. 1, p. 012015, 2024, doi: https://doi.org/10.1088/1755-1315/1415/1/012015. (in English)

[4] T. H. Kramarenko and V. M. Kramarenko, "The use of digital technologies in education in the context of sustainable development of society," *IOP Conf. Ser.: Earth Environ. Sci*., vol. 1415, no. 1, p. 012013, 2024, doi: https://doi.org/10.1088/1755-1315/1415/1/012013. (in English)

[5] S. R. Nath, "Learning losses amid COVID-19 school closures: evidence from Bangladesh," *Irish Educ. Stud.*, 2024, doi: https://doi.org/10.1080/03323315.2024.2441160. (in English)

[6] I. V. Kholoshyn, T. G. Nazarenko, S. V. Mantulenko, O. B. Mazykina, and I. M. Varfolomyeyeva, "Geography of the COVID-19 pandemic in Ukraine and the world: similarities and differences," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1415, no. 1, p. 012042, 2024, doi: https://doi.org/10.1088/1755-1315/1415/1/012042. (in English)

[7] L. Londar and M. Pietsch, "Providing distance education during the war: The experience of Ukraine," *Inf. Technol. Learn. Tools*, vol. 98, no. 6, p. 3151, 2023, doi: https://doi.org/10.33407/itlt.v98i6.5454. (in English)

[8] O. Topuzov, O. Malykhin, N. Aristova, L. Kalinina, and I. Mosiakova, "Learning losses among students of general secondary schools in Ukraine as a consequence of unpredictable global challenges: Challenges and perspectives," Society. *Soc. Integr. Educ. Proc. Int. Sci. Conf.*, vol. 1, pp. 563–572, 2024, doi: https://doi.org/10.17770/sie2024vol1.7915. (in English)

[9] S. Papadakis, S. O. Semerikov, A. M. Striuk, H. M. Kravtsov, M. P. Shyshkina, and M. V. Marienko, "Embracing digital innovation and cloud technologies for transformative learning experiences," in *Proc. 11th Workshop Cloud Technol. Educ. (CTE 2023)*, Kryvyi Rih, Ukraine, Dec. 22, 2023, pp. 1–21. [Online]. Available: https://ceur-ws.org/Vol-3679/paper00.pdf (in English)

[10] A. Artyukhov, A. Simakhova, N. Artyukhova, M. Bojaruniec, and B. Wit, "Information support of e-learning: Ukrainian challenges and cases during the war," *J. Mod. Sci.*, vol. 54, no. 5, pp. 338–354, 2023, doi: https://doi.org/10.13166/jms/176381. (in English)

[11] Z. Blasko, P. da Costa, and S. V. Schnepf, "Learning losses and educational inequalities in Europe: Mapping the potential consequences of the COVID-19 crisis," *J. Eur. Soc. Policy*, vol. 32, no. 4, pp. 361–375, 2022, doi: https://doi.org/10.1177/09589287221091687. (in English)

[12] O. M. Topuzov, Ed., *Diagnostics and compensation of educational losses in general secondary education of Ukraine: methodological recommendations*. Kyiv: Pedahohichna dumka, 2023. [Online]. Available: https://lib.iitta.gov.ua/id/eprint/736974/ (in Ukrainian)

[13] M. Jakubowski, T. Gajderowicz, and H. A. Patrinos, "Global learning loss in student achievement: First estimates using comparable reading scores," *Econ. Lett.*, vol. 232, p. 111313, 2023, doi: https://doi.org/10.1016/j.econlet.2023.111313. (in English)

[14] H. B. Kenayathulla and M. Yemini, "Recovering educational losses post-pandemic: What's next?," *Int. J. Educ. Dev.*, vol. 103, p. 102879, 2023, doi: https://doi.org/10.1016/j.ijedudev.2023.102879. (in English)

[15] V. Ponomarenko and O. Pysarchuk, "Peculiarities of the impact of learning losses on the formation of human capital in Ukraine under martial law," *Econ. Dev.*, vol. 23, no. 1, pp. 38–52, 2024, doi: https://doi.org/10.57111/econ/1.2024.38. (in English)

[16] *War and education: 2 years of full-scale invasion. Report on study results*, Kyiv, 2024. [Online]. Available: https://mon.gov.ua/static-objects/mon/sites/1/news/2024/02/10/Report.War.and.education.Two.years.of.full.scale.invasion.2024.eng-10.02.2024.pdf(in Ukrainian)

[17] M. Szekely, I. Flores-Cecena, F. Hevia, and D. Calderon, "Measuring learning losses from delayed return to school: Evidence from Mexico," *Int. J. Educ. Dev.*, vol. 106, p. 102998, 2024, doi: https://doi.org/10.1016/j.ijedudev.2024.102998. (in English)

[18] M. Velykodna, N. Mishaka, Z. Miroshnyk, and V. Deputatov, "Primary education in wartime: How the Russian invasion affected Ukrainian teachers and the educational process in Kryvyi Rih," *Rev. Rom. Educ. Multidimens.*, vol. 15, no. 1, pp. 285–309, 2023, doi: https://doi.org/10.18662/rrem/15.1/697. (in English)

[19] O. Bilyakovska, N. Mospan, and K. Binytska, "Assessment model of high school students performance: Experience of Ukraine," *Continuing Prof. Educ.: Theory Pract.*, no. 3, pp. 7–18, 2021, doi: https://doi.org/10.28925/1609-8595.2021.3.1. (in English)

[20] S. M. Aslam, A. K. Jilani, J. Sultana, and L. Almutairi, "Feature evaluation of emerging e-learning systems using machine learning: An extensive survey," *IEEE Access*, vol. 9, pp. 69573–69587, 2021, doi: https://doi.org/10.1109/access.2021.3077663. (in English)

[21] L. Shesterova, D. Pyatnytska, S. Synytsya, and T. Synytsya, "Overcoming educational losses of junior class schoolchildren through physical education," *Pedagogical Sciences*, no. 1, pp. 61–65, 2024, doi: https://doi.org/10.33989/2524-2474.2024.1.308723. (in Ukrainian)

[22] V. Matzavela and E. Alepis, "Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments," *Comput. Educ.: Artif. Intell.*, vol. 2, p. 100035, 2021, doi: https://doi.org/10.1016/j.caeai.2021.100035. (in English)

[23] M. C. Gamboa Mora, F. Vivian Mohr, V. Ahumada De La Rosa, S. P. Vera-Monroy, and A. Meja-Camacho, "Predictive model for the classification of university students at risk of academic loss," *Educ. Humanismo*, vol. 26, no. 47, 2024, doi: https://doi.org/10.17081/eduhum.26.47.6379. (in English)

[24] T. Song, X. Luo, and L. Li, "Clustering of basic educational resources and urban resilience development in the central region of China: An empirical study based on POI data," *Reg. Sci. Environ. Econ.*, vol. 1, no. 1, pp. 46–59, 2024, doi: https://doi.org/10.3390/rsee1010004. (in English)
[25] O. Pronina and O. Piatykop, "Predicting students academic performance based on the cluster analysis method," in *ICTERI 2021 Workshops*, vol. 1635, O. Ignatenko, et al., Eds. Cham: Springer, 2022, pp. 292–307, doi: https://doi.org/10.1007/978-3-031-14841-5_19. (in English)
[26] A. A. P. Pamungkas, D. Maryono, and C. W. Budiyanto, "Cluster analysis for student grouping based on index of learning styles," *J. Phys.: Conf. Ser.*, vol. 1808, no. 1, p. 012023, 2021, doi: https://doi.org/10.1088/1742-6596/1808/1/01202. (in English)

Text of the article was accepted by Editorial Team *18.09.2025р.*

КЛАСТЕРНИЙ АНАЛІЗ НАВЧАЛЬНИХ ДАНИХ ДЛЯ ОЦІНЮВАННЯ ТА ВІЗУАЛІЗАЦІЇ ОСВІТНІХ ВТРАТ У ШКІЛЬНІЙ ОСВІТІ: ПРОГРАМНА РЕАЛІЗАЦІЯ

Ольга Проніна
кандидат технічних наук, доцент кафедри комп'ютерних наук,
ДВНЗ «Приазовський державний технічний університет», м. Дніпро, Україна
ORCID ID 0000−0001−7085−8027
pronina_o_i@pstu.edu

Олена Броднюк
магістр
ДВНЗ «Приазовський державний технічний університет», м. Дніпро, Україна
lbrodniuk@ukr.net

Олена П'ятикоп
кандидат технічних наук, доцент кафедри комп'ютерних наук
ДВНЗ «Приазовський державний технічний університет», м. Дніпро, Україна
ORCID ID 0000−0002−7731−3051
piatykop_o_ye@pstu.edu

Ірина Федосова
доктор педагогічних наук, професор кафедри комп'ютерних наук
ДВНЗ «Приазовський державний технічний університет», м. Дніпро, Україна
ORCID ID 0000-0003-3923-8270
fedosova_i_v@pstu.edu

Анотація. Стаття присвячена актуальній проблемі освітніх втрат, що виникли внаслідок пандемії COVID-19 та воєнних дій в Україні. У статті розглядаються поняття освітніх втрат, передумови та причини їх виникнення, а також категорії втрат. Також аналізується досвід інших науковців щодо вимірювання та компенсації освітніх втрат. Окремо аналізуються втрати в навчанні, зниження рівня академічних досягнень учнів. Автори пропонують комплексний підхід до вимірювання та компенсації освітніх втрат з використанням методів кластерного аналізу. Застосовано алгоритми k-середніх та ієрархічної кластеризації, що дали змогу визначити групи учнів із подібними навчальними досягненнями. Для обробки та аналізу даних використано Python із бібліотеками Pandas, Matplotlib і Scikit-learn; реалізовано графічний інтерфейс для візуалізації результатів, що полегшує роботу вчителя. У статті представлені експериментальні дослідження щодо групування учнів 5-6 класів у кластери, а також розрахунки втрат, отримані за допомогою розробленого програмного забезпечення. Результати підтвердили наявність стійкої кореляції між досягненнями з різних предметів і дали змогу обчислити рівень індивідуальних освітніх втрат.
Запропонований підхід створює умови для персоналізації навчання, своєчасної діагностики прогалин та планування компенсаторних заходів, включно з диференційованими завданнями, груповим репетиторством і додатковою психологічною підтримкою. Зроблено висновок, що автоматизація аналізу освітніх даних і використання кластерного аналізу суттєво підвищують ефективність подолання освітніх втрат та сприяють забезпеченню рівного доступу до якісної освіти навіть у кризових умовах.

Це дозволить учителям краще організовувати компенсацію освітніх витрат, адаптувати навчальні програми відповідно до індивідуальних потреб учнів.

Ключові слова: освітні втрати; дистанційне навчання; кластерний аналіз; індивідуалізація навчання; k-means; ієрархічна кластеризація

This work is licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.