Analyzing Students' Online Activity to Enhance Education Quality and Boost University Digital Security

Valerii Lakhno[1], Nurzhamal Oshanova[2], Jamilya Akhmetova[3], Nurgazy Kurbaniyazov[4], and Miroslav Lakhno[1]

[1] *National University of Life and Environmental Sciences of Ukraine, 6a Heroyiv oborony str., Kyiv, 03041, Ukraine*
[2] *Kazakh National Pedagogical University named after Abay, 13 Dostyk ave., Almaty, 050010, Kazakhstan*
[3] *Sh. Yesenov Caspian University of technology and engineering, 32 Microdistrict, Aktau, 130000, Kazakhstan.*
[4] *Kazakh National University named after Al-Farabi, 71 Al-Farabi ave., Almaty, 050038, Kazakhstan*

Abstract

This article proposes an algorithm for a Decision Support System (DSS) that helps to improve the quality of education and the level of security of the Digital Educational Environment of Universities (DEEU), based on the analysis of users' Digital Tracks (DTs). The algorithm is based on the matrix factorization of users' DT. In contrast to known solutions, the proposed solution allows us to level the problem of developing the competency profile of students and employees, primarily in matters of acquiring Information Security (IS) skills. This will contribute to increasing the level of security of the DEEU in general. It is shown that mastering new competencies related to information security issues expands the profile not only of students, but also of university staff, and ultimately contributes to the formation of future specialists who have a proactive position in information security issues and are capable of self-organization in this subject area. The implementation of artificial intelligence methods in such DSS can be realized based on machine learning, using, for example, the matrix factorization of DTs.

Keywords [1]

Learning quality, digital tracks, matrix factorization, machine learning, information security.

1. Introduction

The creation of a DEEU is a key element of the modern educational system in any developed country. Publications [1–3] show that with the development of Information Technologies (IT), a high-quality DEEU contributes significantly to the improvement of the quality of education. The creation of a centralized information management system has several important reasons: relevance and accessibility of information, interactive training, flexibility and mobility, individualization of training, development of digital skills, saving resources of educational institutions, etc. [4, 5]. During training, students leave various so-called DTs. Here is just a small list of such DTs left by students during their studies and interactions with the DEEU [6–8]: email; online learning platforms; social media; digital files; internet searches; etc. These DTs can be useful for students, teachers, and university administrators for tracking educational progress, assessing work, communicating, and analyzing data to improve the quality of the educational process. No less important, however, is the task of ensuring the confidentiality and protection of these DTs and the personal data of students and teachers. The latter is because in many cases the DT of DEEU users contain personal information. Therefore, the information contained in the DT may be

CPITS-2024: Cybersecurity Providing in Information and Telecommunication Systems, February 28, 2024, Kyiv, Ukraine
EMAIL: lva964@nubip.edu.ua (V. Lakhno); uruytihonov1@gmail.com (Y. Tikhonov); sab_nazym@mail.ru (N. Sabitova); taimuratova@mail.ru. (L. Taimuratova); mereke.zhumadilova@yu.edu.kz (M. Zhumadilova)
ORCID: 0000-0001-9695-4543 (V. Lakhno); 0009-0003-9365-618X (Y. Tikhonov); 0009-0001-2531-5653 (N. Sabitova); 0000-0002-1692-4350 (L. Taimuratova); 0000-0002-2974-0227 (M. Zhumadilova)

subject to cyber-attacks or misuse. In the context of the globalization of education [9], universities should adapt their IS policies to protect data in the DT to the maximum extent possible, and inform students of the institution's strategy for protecting user data, its IS policy, and specific practices for protecting their DT [10–12].

Information security of the DEEU is a complex system that provides for the protection of the information space available in an educational institution. Such a system or systems make it impossible to damage or steal the personal data of participants in the educational process, as well as information that has financial, intellectual value, etc. Ensuring the effective functioning of the information security system of DEEU requires the expenditure of certain financial resources within the framework of the data protection strategy chosen by the educational institution. When developing such a strategy, it is advisable to take into account factors of the external and internal environment, since achieving an optimal result can only be achieved if a balance is found between the available capabilities and the desired results. These results also include the integral goal of improving the quality of education by leveraging the potential of all forms of organizing the educational process and developing the infrastructure of universities in the context of digital transformation. In such a situation, for the management of an educational institution and personnel responsible for information security policy, a context-driven approach to intelligent decision support for ensuring information security of the centralized information security system based on the analysis of users' central systems may become in demand.

All of the above predetermined our interest in research in this area.

The purpose of the study is to develop an algorithm for a decision support system that, based on the analysis of DT, contributes to improving the quality of education and the degree of security of the DEEU.

2. Methods and Models

When analyzing DT in a DEEU, administrators, and information security specialists are most often concerned with analyzing log files. In addition, log files and DT are related in the context of analyzing digital evidence of user actions. Here are just a few examples of the relationship between log files and users' DT in the CSO: information storage in the DEEU; event recovery; data analysis; etc. Thus, log files and DT are interrelated, since log files contain information about events and actions that can be analyzed to identify, locate, and interpret DT in the context of analyzing digital evidence of the activities of users of DT, both in the context of improving the quality of the educational process and in the context of IS DEEU. Log files are independent characteristics of a user's work on the university network. They contain information about system logins, resource usage, errors, network activity, and other events in the data center.

However, to fully understand the context of a user's experience, log files typically require analysis and interpretation by other tools. Contextual characteristics of a user's experience in the DEEU can include information about time, location, applications used, and other factors that may be related to the user's specific situation or task in the DEEU. These characteristics can be extracted from log files. However, additional analysis and context is usually required. Such analysis can be implemented using specialized software, such as Splunk, ELK Stack (a software stack that includes Elasticsearch, Logstash, and Kibana), etc. [13, 14].

The level of security of the DEEU can be increased in particular by using decision support systems, artificial intelligence, and machine learning methods.

In the context of the research objectives, it is necessary to develop an algorithm for DSS that, based on the analysis of the DT, helps to improve the quality of education and the level of security of the DEEU. In the approach proposed below, artificial intelligence is expressed through machine learning techniques using matrix factorization.

We believe that the DEEU has many users: and tasks related to information security $U = \{u_1, ..., u_n\}$ both educational tasks $E = \{e_1, ..., e_m\}$ and tasks related to IS DEEU—$S = \{s_1, ..., s_k\}$. Then administrators of the DEEU have access to matrices containing, for example:

- Ratings are given by users based on their priority for educational tasks—$ME_{n \times m}$.

- And, also characterizing the user from the point of view of compliance with information security rules when working in the DEEU—$MS_{n \times k}$.

In the matrix $ME_{n \times m}$, a certain number is put in place of $me_{ij}(i \in 1, \ldots, n; j \in 1, \ldots, m)$ if the user of the DEEU (u_i) evaluates the task (e_j) based on his priorities, and remains empty otherwise. Data is taken based on user DT from Moodle, Blackboard, Canvas, Google Classroom, etc.

To fill out the second matrix according to the criteria for safe behavior in the DEEU, we will identify the following types of users according to the level of their competence in information security issues:

Knowledgeable users. This group includes users who are well aware of the information security risks in the university network or the DEEU as a whole. Such users take active measures to ensure the information security of their data and accounts in the DEEU. Such users punctually follow recommendations for creating complex passwords, regularly update software, do not open suspicious links or attachments in emails, and use reliable antivirus software.

Careless users. This group includes users who do not pay due attention to information security measures. Consequently, they are quite vulnerable to attacks, both external and internal. Users of this group typically use weak passwords and repeat them for different accounts. These users, usually, ignore suspicious activity on the network and do not comply with measures to protect their data in the DEEU.

Unaware users. This group includes users who do not have a sufficient level of knowledge about information security measures when working on the network. They may not be aware of the risks associated with opening suspicious links, and they may not be familiar with the rules for using public Wi-Fi networks in the DEEU. Such users install untrustworthy software without fear and often transmit confidential data through unsecured communication channels.

Indifferent users. This group includes users who do not show any interest in information security issues on the university network and who do not adhere to basic information security measures.

Irresponsible users. This group includes users of the DEEU who violate the rules and policies of information security on the university network. They may attempt to gain unauthorized access to the DEEU, distribute malware, violate data confidentiality, or engage in fraudulent activities within the DEEU.

The above categorization of user types is rather arbitrary. As noted in [15], there are no clear boundaries between the mentioned categories of users. As they gain knowledge, for example through relevant courses in university curricula, users may move from one type to another, recognizing the importance of online information security and taking appropriate measures to protect their data and accounts.

Then in the matrix $MS_{n \times k}$, a certain number is placed in place $ms_{ij}(i \in 1, \ldots, n; j \in 1, \ldots, k)$ if the user of the DEEU (u_i) is assigned to a certain group (s_j), based on the style of his behavior in the DEEU in the context of compliance with information security rules.

In effect, this matrix displays data related to the information security competencies of students and employees. In a digitized form, such a matrix may contain, for example, the behavior style of a student or employee in information security matters. Such data has been obtained based on the analysis of the DT, e.g. using the methodology given in [16, 17]. Otherwise, i.e. when the IS style assessment is not performed, the space remains empty.

It is required to find vectors $(\widehat{me}_{\iota})$, $(\widehat{ms}_{\iota})$ containing data about:

1. In the context of the formation of an individual educational trajectory of already known assessments of the user (u_i), i.e. $(\widehat{me}_{\iota})$. Also estimated estimates—$(\widehat{me_{\iota J}})$.
2. In the context of developing skills for safe work in the DEEU of already known skills, for example, based on the DT or testing results, i.e. $(\widehat{ms}_{\iota})$. Also estimated assessments, after the acquisition, of the relevant competencies in information security—$(\widehat{ms_{\iota J}})$.

Since one of the research objectives was to develop an algorithm for DSS, Matrix Factorization (MF) was used as a machine learning method. MF meant the decomposition of the original matrix into the product of two matrices of small rank [18, 19]. Accordingly, the interaction of users with an object will be modeled as a scalar product of vectors of representation of users and objects in a factor space relating, for example, to the competencies of students in information security issues. Note that factorization models have proven themselves well when working with highly sparse matrices [18, 19]. This is because MF allows you to extract hidden dependencies based on the analysis of users' DT in the DEEU and make predictions based on large volumes of information circulating in any educational institution.

In the context of machine learning, DSS MF can be used, for example, for the tasks of developing recommendations related both to improving the quality of the educational process as a whole and to individual competencies of students and employees, for example in information security.

Since working with matrices is similar, in the context of this study we consider only the algorithm for working with the matrix of student assessments in the DEEU, see Table 1 and Fig. 1.

Let us present the assessment matrix $ME_{n\times m}$ as a product of two matrices:

Matrix $A_{n\times w}$, which includes a numerical description of hidden (latent) characteristics of users (for example, behavioral patterns: regularity of activity in the DEEU, frequency and time of entry into the DEEU, typical activity intervals; content consumption; access level, etc., frequency of erroneous logins, attempts to access prohibited resources, etc., as well as explicit ones (course, age, average grades, etc.).

Matrix $B_{w\times u}$, which characterizes educational tasks, for example, the priority of courses in IT and/or information security for the formation of an individual educational trajectory.

We fill in random variables based on the law of uniform distribution on the interval $\left[0;\ \sqrt{\max\{me_{ij}\}/k}\,\right]$ of the latent characteristics for, respectively, the matrices *A* and *B*.

Then solve the minimization problem using the dependence (1):

$$\arg\min\left\|ME-\hat{ME}\right\|+\alpha\|B\|+\beta\|A\|, \qquad (1)$$

where *ME* is the matrix that is obtained as a result of approximation from the matrices *A* and *B*, *α*, *β* which are algorithm parameters.

At each step of the iterative algorithm, error minimization will include the following steps presented in Table 1 and Fig. 1.

Table 1

Steps of the algorithm for finding a solution for minimizing errors in machine learning based on matrix factorization of data from the analysis of the DT of users of the DEEU

Step number	Description of action	Mathematical interpretation and decoding of parameters
1	Finding the matrix B	
2	Identify the error δ	$\delta=\left|me-\hat{me}\right|,\ j\in 1,\dots,n$
3	Finding new values A_{ir}	$A_{ir}=A_{ir}-v\left(\delta B_{rj}^{E}+\lambda A_{ir}\right),$ where $r\in 1,\dots,k,\ \lambda -$ regulatory parameter. $v -$ learning rate
4	Finding the matrix A	
5	Identify the error δ	$\delta=\left|me-\hat{me}\right|,\ j\in 1,\dots,m$
6	Finding new values B_{rj}	$B_{rj}=B_{rj}-v\left(\delta A_{ir}^{E}+\lambda B_{rj}\right)$

As the volume of data obtained increases, and therefore the data sparsity decreases, an increase in the number of iterations may be required. Control of the number of iterations can be automated. Currently, the development of appropriate software is underway, with the help of which, after factorization and comparison of the accuracy of predictions with past results, it will be possible to develop, using DSS, to improve the quality of the educational process, and, in particular, competencies in information security. If the prediction accuracy decreases, the number of iterations should be increased. Otherwise, the number of iterations, see Fig. 1, will not change.

As an alternative, the acceptable factorization precision can be specified. If this accuracy is achieved, then the algorithm shown in Fig. 1 stops.

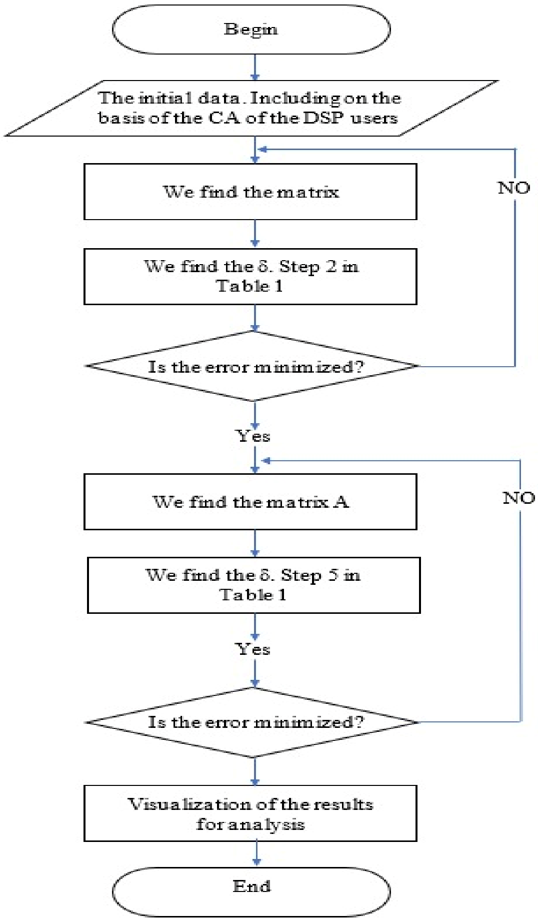
Begin
The initial data. Including on the basis of the CA of the DSP users
We find the matrix
We find the δ. Step 2 in Table 1
Is the error minimized?
NO
Yes
We find the matrix A
We find the δ. Step 5 in Table 1
Yes
Is the error minimized?
NO
Visualization of the results for analysis
End

Figure 1: Block diagram of the matrix factorization algorithm based on the analysis of data from users' DT in the DEEU

As indicated above, working with the second matrix, concerning the categorization of user types on information security issues, is similar.

3. Conclusions

As part of the study, an algorithm was proposed for a DSS that helps improve the quality of education and the degree of security of the DEEU based on the analysis of users' DT. The algorithm is based on matrix factorization of users' DT in the DEEU. The proposed solution allows us to mitigate the problem of developing the competency profile of students and employees, primarily in matters of acquiring IS skills. This, in our opinion, generally contributes to increasing the degree of security of the central communication system. Mastering new competencies related to information security issues expands the profile of not only students but also university employees. This, ultimately, contributes to the formation of future specialists who have a proactive position in information security issues and are capable of self-organization in this subject area. It is shown that the implementation of artificial intelligence methods in such DSS can be realized based on machine learning using matrix factorization. The implementation of such intellectual tools in such DSS will make it possible to qualitatively select educational material, in particular, on information security issues. Such material will be of interest to students since it is focused on their characteristics and will involve them in the learning process as much as possible.

Acknowledgments

The work was carried out within the framework of grant research IRN AR 19678846 "Increasing the efficiency of hybrid and distance forms of organizing the educational process based on the development of university infrastructure in the context of digital transformation" (Republic of Kazakhstan).

References

[1] R. Muydinovich, M. Valentinovna, M. Xabibjonqizi, The Role of Information Technology in Modern Methods in the System of Higher Education, Int. J. Early Child. Spec. Educ. 14(7) (2022).

[2] V. Kashuba, I. Asaulyuk, A. Diachenko, A Modern View on the Use of Information Technologies in the Process of Physical Education of Student Youth, J. Educ. Health Sport 7(2) (2017) 765–775.

[3] A. Haleem, et al., Understanding the Role of Digital Technologies in Education: A review, Sustainable Operations and Computers 3 (2022) 275–285. doi: 10.1016/j.susoc.2022.05.004.

[4] V. Buriachok, et al., Implementation of Active Cybersecurity Education in Ukrainian Higher School, Information Technology for Education, Science, and Technics, vol. 178 (2023) 533–551. doi:10.1007/978-3-031-35467-0_32.

[5] V. Buriachok, V. Sokolov, Implementation of Active Learning in the

Master's Program on Cybersecurity, Advances in Computer Science for Engineering and Education II, vol. 938 (2020) 610-624. doi: 10.1007/978-3-030-16621-2_57.

[6] A. Vaccari, et al., Towards the SocioScope: An Information System for the Study of Social Dynamics Through Digital Traces, 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (2009) 52–61. doi: 10.1145/1653771.1653782.

[7] V. Mantulenko, Prospects of Digital Footprints Use in the Higher Education, Current Achievements, Challenges and Digital Chances of Knowledge Based Economy, LNNS 133 (2021) 581–589. doi: 10.1007/978-3-030-47458-4_67.

[8] N. Morze, O. Kuzminska, M. Mazorchuk, (2019). Attitude to the Digital Learning Environment in Ukrainian Universities, In: ICT in Education, Research, and Industrial Applications Vol. 2393 (2019) 53–67.

[9] A. Goodwin, Globalization, Global Mindsets and Teacher Education, Action Teacher Educ. 42(1) (2020) 6–18. doi: 10.1080/01626620.2019.1700848.

[10] P. Skladannyi, et al., Improving the Security Policy of the Distance Learning System based on the Zero Trust Concept, in: Cybersecurity Providing in Information and Telecommunication Systems, vol. 3421 (2023) 97–106.

[11] R. Marusenko, V. Sokolov, P. Skladannyi, Social Engineering Penetration Testing in Higher Education Institutions, Advances in Computer Science for Engineering and Education VI, vol. 181 (2023) 1132–1147. doi: 10.1007/978-3-031-36118-0_96.

[12] R. Marusenko, V. Sokolov, V. Buriachok, Experimental Evaluation of Phishing Attack on High School Students, Advances in Computer Science for Engineering and Education III, vol. 1247 (2020) 668–680. doi:10.1007/978-3-030-55506-1_59.

[13] M. Aarthi, Using Users Profiling to Identifying an Attacks, Turkish J. Comput. Math. Educ. 12(7) (2021) 795–802

[14] K. Subramanian, Introducing the Splunk Platform, Practical Splunk Search Processing Language (2020) 1–38. doi: 10.1007/978-1-4842-6276-4_1.

[15] K. Shu, et al., The Role of User Profiles for Fake News Detection, International Conference on Advances in Social Networks Analysis and Mining (2019) 436–439. doi: 10.1145/3341161.33429 27.

[16] I. Sen, et al., A Total Error Framework for Digital Traces of Human Behavior on Online Platforms, Public Opinion Quarterly 85(S1) (2021) 399–422. doi: 10.1093/poq/nfab018.

[17] R. Coulter, et al., Data-Driven Cyber Security in Perspective—Intelligent Traffic Analysis, IEEE Transactions on Cybernetics, 50(7) (2019) 3081–3093. doi: 10.1109/tcyb.2019.2940940.

[18] C. Lei, et al., A Service Recommendation Algorithm with the Transfer Learning Based Matrix Factorization to Improve Cloud Security, Inf. Sci. 513 (2020) 98–111. doi: 10.1016/j.ins.2019.10.004.

[19] X. Zheng, et al., A Matrix Factorization Recommendation System-Based Local Differential Privacy for Protecting Users' Sensitive Data, IEEE Transactions on Computational Social Systems 10(3) (2022) 1189–1198. doi: 10.1109/tcss.2022.3170691.