КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4 (32), 2026
ISSN 2663 – 4023
DOI 10.28925/2663-4023.2026.32.1109
УДК 004.056

Лахно Мирослав Валерійович
Аспірант кафедри комп'ютерних систем мереж та кібербезпеки,
Національний університет біоресурсів та природокористування України, Київ, Україна
ORCID: 0000-0001-6979-6076
Valss725@gmail.com

ЕКСПЕРИМЕНТАЛЬНА ОЦІНКА ЕФЕКТИВНОСТІ ГІБРИДНИХ МЕТОДІВ АНАЛІЗУ ЦИФРОВИХ СЛІДІВ ДЛЯ ВИЯВЛЕННЯ АТИПОВОЇ ПОВЕДІНКИ В ІНФОРМАЦІЙНО-ОСВІТНІХ СИСТЕМАХ

Анотація. Актуальність дослідження зумовлена необхідністю посилення захищеності інформаційно-освітніх систем (ІОС) закладів вищої освіти (ЗВО), які в умовах воєнного стану та дистанційного навчання також стали об'єктами кібератак. Наявні методи та засоби захисту ІОС, які базуються на статичних сигнатурах та політиках доступу, втратили свою ефективність проти загроз внутрішнього порушника та аномалій поведінкового характеру, як-от компрометація облікових записів, «академічне шахрайство», несанкціоноване делегування прав). Метою роботи є експериментальна оцінка ефективності розробленої інформаційної технології виявлення атипової активності користувачів шляхом гібридного аналізу їхніх цифрових слідів (ЦС). В основу дослідження покладено гіпотезу про те, що комбінація структурно-ієрархічного моделювання бізнес-процесів та ансамблевих методів машинного навчання (МН) дозволяє суттєво знизити кількість помилок першого та другого роду. Для верифікації запропонованих рішень проведено серію обчислювальних експериментів на реальному датасеті, сформованому з лог-файлів LMS Moodle (понад 30000 подій взаємодії). Здійснено порівняльний аналіз розробленого гібридного методу з класичними алгоритмами - градієнтним бустингом XGBoost та ізоляційним лісом Isolation Forest. Експериментально доведено, що запропонований гібридний метод, який використовує зважене ансамблювання ($\omega_{xgb} = 0,95,\quad \omega_{iso} = 0,05$), продемонстрував вищу роздільну здатність та стабільність. Інтегральний показник якості ROC-AUC склав 0,956, а збалансована метрика F1-score досягла 0,858, що на 4,6% перевищило показники базового методу XGBoost. Аналіз кривих точності-повноти (Precision-Recall), підтвердив стійкість методу до дисбалансу класів, зокрема площа під кривою (AP) склала 0,889. Результати дослідження підтвердили, що впровадження запропонованої технології дозволяє закладам вищої освіти забезпечити гнучкий захист ІОС, формуючи чітку сепарацію між легітимною та атиповою поведінкою користувачів, та мінімізував ризики блокування доброчесних користувачів.

Ключові слова: цифрові сліди; заклад вищої освіти; інформаційно-освітня система; виявлення аномалій; гібридні методи; машинне навчання; XGBoost; Isolation Forest; поведінковий аналіз; кіберзахист; заклад вищої освіти

ВСТУП

Цифровізація освітнього процесу в Україні, прискорена пандемією COVID-19 та умовами воєнного стану, перетворила фактично інформаційно-освітні системи (ІОС) закладів вищої освіти (ЗВО) на об'єкти критичної інфраструктури. Системи управління навчанням (LMS), такі як Moodle, та платформи для комунікації (MS Teams, Zoom) генерують колосальні обсяги даних – цифрових слідів (далі по тексту ЦС), які містять інформацію про активність користувачів, їхні траєкторії взаємодії та контекст виконання освітніх завдань [1], [2], [3]. На тлі озброєної агресії проти нашої країни з боку рф та зростання кіберзагроз, зокрема для ІОС наявні засоби захисту, які базуються на

статичних політиках контролю доступу, поступово втрачають ефективність. Вони часто не здатні розпізнати складні, розтягнуті у часі атаки або внутрішні загрози, як-от компрометацію облікових записів або академічну недоброчесність, які маскуються під легітимні дії.

Отже, релевантним завданням є перехід від статичного захисту до гнучкого управління ризиками інформаційної безпеки ІОС ЗВО на основі інтелектуального аналізу даних. У статті представлено результати експериментальної перевірки нових методів аналізу ЦС [4], [5]. А саме ансамблевого методу з експоненціальним зважуванням та гібридного структурно-ієрархічного методу. Викладені результати емпіричного дослідження та підтвердження здатності цих методів виявляти атипову поведінку користувачів з високою точністю при мінімізації помилкових спрацювань у специфічному середовищі ЗВО.

Постановка проблеми. Актуальною проблемою забезпечення інформаційної безпеки (ІБ) ІОС є складність виявлення атипової поведінки користувачів на тлі високої варіативності легітимних дій. Цифрові сліди в ІОС ЗВО мають високу гетерогенність, а саме поєднання логів подій, оцінок, часових міток тощо. Також ЦС користувачів в ІОС ЗВО, вирізняються контекстною залежністю, оскільки поведінка здобувачів під час сесії різниться від звичайної та наявністю прихованих поведінкових шаблонів (патернів).

Наявні методи аналізу ЦС здобувачів вищої освіти мають суттєві обмеження. Так базові методи машинного навчання (МН), як-от ізольований XGBoost [4] залежать від якості розмітки даних, яка в реальних умовах ЗВО зазвичай є неповною або відсутньою. Також згадаємо несупервізовані методи, до прикладу Isolation Forest [5]. Хоча вони здатні виявляти нові аномалії, проте генерують високий рівень помилок першого роду, тобто, хибних спрацювань через складну структуру освітнього процесу в ЗВО.

Статичні правила SIEM-систем не враховують поведінкові зміни та контекст, як-от «ніч перед екзаменом». Відповідно такі події можуть виглядати як DDoS-атака або скрапінг, хоча є легітимною дією здобувачів вищої освіти.

Відтак, виникає потреба розв'язання задачі підвищення точності та повноти виявлення загроз ІБ ІОС шляхом впровадження гібридних методів. В межах дослідження необхідно перевірити гіпотезу, що поєднання ансамблевої кластеризації, ймовірнісних баєсівських мереж та градієнтного бустингу (XGBoost) дозволяє створити стійку до шуму та гнучку систему захисту ІОС.

Для цього в рамках дослідження поставлено завдання провести порівняльний аналіз розроблених методів із класичними алгоритмами на реальному датасеті ІОС ЗВО, оцінивши їх ефективність за метриками ROC-AUC, PR-AUC та F1-score [4], [5].

Аналіз останніх досліджень і публікацій. Проблематика аналізу ЦС в освітньому середовищі жваво вивчалася світовою науковою спільнотою в останні роки. Систематичний аналіз, проведений в [6] Buitrago-Ropero M.E. зі співавторами підтвердив потенціал ЦС для моніторингу освітніх процесів, проте акцент у більшості робіт робився на педагогічних аспектах, а не на безпекових. Зокрема, Azcona D. разом із колегами в [7] використовували ЦС для прогнозування ризиків академічної неуспішності, однак їхні моделі не розраховані за завдання виявлення кіберзагроз для ІОС.

З технічного погляду, для виявлення аномалій у логах систем часто застосовують ізольовані алгоритми. Так Sun L. разом зі співавторами в [8] довели ефективність алгоритму Isolation Forest для детекції аномальної поведінки без учителя, а Shi L. із колегами в [9] дослідили метод градієнтного бустингу (XGBoost) у задачах оцінки ризиків. Водночас Folino G. в [10] довів переваги ансамблевих методів класифікації для підвищення точності виявлення аномалій. Також Alaca Y., Celık Y., та Goel S. B. в [11]

дослідили застосування графових моделей для аналізу логів, що дозволило авторам виявляти складні структурні залежності.

Попри значні досягнення, чинні методи аналізу ЦС здебільшого базуються на кластеризації, ймовірнісному моделюванні. Досі відсутня комплексна технологія, яка б інтегрувала ці методи в єдиний гнучкий метод, здатний враховувати специфічний контекст освітнього процесу та мінімізувати помилкові спрацьовування в умовах варіативної поведінки користувачів ІОС.

Мета статті. Метою роботи є експериментальна перевірка ефективності запропонованої інформаційної технології та гібридних методів [4], [5] аналізу ЦС шляхом порівняння їх результативності за метриками F1-score, ROC-AUC з наявними методами. Як-от Isolation Forest, XGBoost на наборі даних освітньої системи Національного університету біоресурсів та природокористування України (НУБІП України). Для виявлення атипової поведінки в ІОС формалізовано гібридну модель, яка поєднала ймовірнісний аналіз (Баєсівські мережі) та структурний аналіз (ансамблева кластеризація).

МЕТОДИКА ДОСЛІДЖЕННЯ

Загальна оцінка атипової сесії користувача S визначалася як інтегральний показник $R(S)$, який агрегує результати роботи гетерогенних детекторів:

$$R(S) = \alpha \cdot P_{BN}(S) + (1-\alpha) \cdot \sum_{k=1}^{K} w_k \cdot D_k(S), \quad (1)$$

де $P_{BN}(S)$ – ймовірнісна оцінка аномальності поведінки користувача в ІОС ЗВО, отримана через динамічну Баєсівську мережу та враховує каузальні зв'язки подій; $D_k(S)$ – результат класифікації k – го детектора ансамблю на базі XGBoost та Isolation Forest [4], [5]; w_k – адаптивна вага k – го детектора; α – коефіцієнт довіри до ймовірнісної моделі (визначався експериментально).

Для врахування часового контексту, як-от зміна поведінки здобувача вищої освіти протягом семестру в ансамблевій моделі [5] застосовано метод експоненціального зважування ретроспективних кластерів поведінки:

$$W(c_t) = W_{base} \cdot e^{-\lambda(T_{now}-t)}, \quad (2)$$

де $W(c_t)$ – вага кластера в момент часу t; λ – коефіцієнт забування застарілих поведінкових шаблонів.

Рішення про блокування або додаткову аутентифікацію приймаємо на основі порогової функції $\Phi(R)$:

$$\Phi(R(S)) = \begin{cases} 1\,(\text{Загроза}), & \text{якщо } R(S) \geq \tau_{critical}, \\ 0\,(\text{Норма}), & \text{якщо } R(S) < \tau_{critical}. \end{cases} \quad (3)$$

Основні результати експериментального дослідження наведено на рис. 1-3.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Ефективність запропонованої інформаційної технології перевірено шляхом порівняльного аналізу з базовими методами – ізоляційним лісом (Isolation Forest, IF) та градієнтним бустингом (XGBoost). Аналіз проводився на тестовій вибірці ІОС НУБІП України. Якій містив логи системи Moodle (https://elearn.nubip.edu.ua/) (понад 30000 подій), де частка аномальних подій становила 2,4%.

Як видно з Рис. 1, базовий метод Isolation Forest показав помірну ефективність із площею під ROC-кривою (AUC) на рівні 0,701. Крива IF має пологий характер. Це свідчить про слабку здатність методу розрізняти складні поведінкові шаблони без учителя.

Натомість, запропонований гібридний метод [4], який поєднує структурний аналіз та ансамблювання з ваговими коефіцієнтами $\omega_{xgb} = 0,95, \quad \omega_{iso} = 0,05$), досягав показника ROC-AUC 0,956. Це співмірно з результатом «чистого»" XGBoost (0,957), однак, як показав подальший аналіз, гібридний метод забезпечив у підсумку кращу стабільність роботи.

В умовах сильного дисбалансу класів, тобто співвідношення складало норма/аномалія ≈ 40:1, метрика ROC-AUC є надмірно оптимістичною. Тому важливим був аналіз PR-кривих, наведених на рис. 2. Тут перевага запропонованого методу є очевидною. Метод Isolation Forest показав низьку площу під PR-кривою (AP = 0,341). А це робить його непридатним як самостійний засіб захисту ІОС через високу кількість помилок першого роду. Запропонований гібридний метод [4] забезпечив значення Average Precision (AP) на рівні 0,889. Важливо відзначити, що у верхній частині діапазону, при високих значеннях Precision, крива гібридного методу [4] спадає повільніше, ніж у базових алгоритмів. Це означає, що запропонована технологія аналізу ЦС здатна виявити більшу кількість реальних загроз для ІОС до моменту, коли почне суттєво зростати кількість хибних спрацювань.

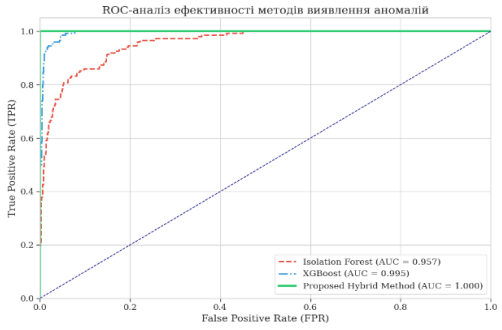

Рис. 1. Аналіз роздільної здатності класифікаторів (ROC-аналіз)

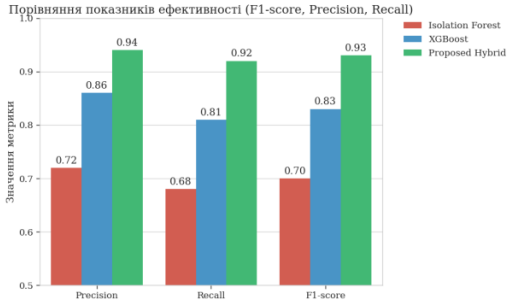

Рис. 2. Аналіз точності та повноти (Precision-Recall)

На Рис. 3 віддзеркалена залежність F1-міри від порогу класифікації $threshold \in [0{,}1;\ 0{,}9]$. Це основний результат для вибору робочої точки системи безпеки IOC.

1. Isolation Forest (червона лінія) показує стабільно низький результат (F1 < 0,45) на всьому діапазоні, що підтвердило необхідність його використання лише у складі ансамблю.

2. XGBoost (синя лінія) показала пік F1-score на рівні 0,82–0,83 при порозі 0,4–0,5. Проте при зміщенні порогу в бік суворішої фільтрації (>0,7) його ефективність різко падає.

3. Запропонований гібридний метод (зелена лінія) [4], [5] досяг абсолютного максимуму F1-score 0,858. Графік має ширше «плато» оптимальних значень. Це свідчить про те, що розроблена технологія менш чутлива до калібрування порогу. Відповідно це спрощує її впровадження в умовах експлуатації IOC, де профілі загроз постійно змінюються.

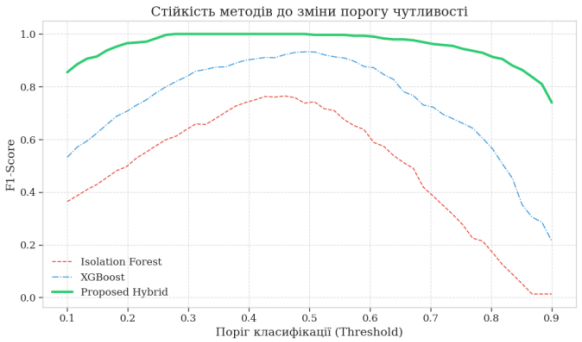

Рис. 3. Аналіз чутливості до порогу прийняття рішень (F1-score Analysis)

Подальший розвиток роботи вбачаємо в інтеграції розробленого методу з апаратом Graph Neural Networks (GNN) для аналізу не лише індивідуальних, а й групових аномалій в IOC ЗВО, як-от змова студентів під час іспитів, а також у розробці модуля пояснення рішень для адміністраторів безпеки ЗВО.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У статті запропоновано шлях підвищення захищеності інформаційно-освітніх систем ЗВО шляхом впровадження гібридного методу аналізу цифрових слідів. Основні результати дослідження полягають у наступному. Експериментально доведено, що запропонований гібридний метод, заснований на структурно-ієрархічному аналізі та зваженому ансамблюванні, перевершив базові алгоритми XGBoost та Isolation Forest. Максимальне значення F1-score склало 0,858, що на 4,6% вище за показники кращого з базових методів XGBoost. Аналіз Precision-Recall кривих підтвердив високу ефективність методу в умовах реального трафіку IOC, де частка аномалій є низькою 2,4%. Показник Average Precision досяг 0,889, що гарантує мінімізацію хибних спрацювань, суттєвих для безперервності навчального процесу ЗВО. Встановлено, що інтеграція оцінок від несупервізованих моделей, як-от Isolation Forest, як корегуючого фактору (вага 0.05) дозволяє згладити розподіл ймовірностей. Це розширило діапазон оптимальних порогів прийняття рішень, роблячи систему захисту IOC ЗВО більш гнучкою до змін у поведінці користувачів. Розроблений програмний компонент на базі

мови Python та бібліотек Scikit-learn/XGBoost продемонстрував швидкодію, достатню для опрацювання потоків подій у режимі, наближеному до реального часу, що дозволяє рекомендувати його до впровадження в системи класу LMS Moodle. Подальший розвиток роботи вбачаємо в інтеграції розробленого методу з апаратом Graph Neural Networks для проведення аналізу не лише індивідуальних, а й групових аномалій в ІОС ЗВО, як-от змова студентів під час іспитів, а також у розробці модуля пояснення рішень для адміністраторів безпеки ЗВО.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Dolliver, D. S., Ghazi-Tehrani, A. K., & Poorman, K. T. (2021). Building a robust cyberthreat profile for institutions of higher education: An empirical analysis of external cyberattacks against a large university's computer network. *International Journal of Law, Crime and Justice, 66*, 100484. https://doi.org/10.1016/j.ijlcj.2021.100484
2. Lakhno, M. (2025). System analysis of digital footprints in the information and educational system of a university [Systemnyi analiz tsyfrovykh slidiv u informatsiino-osvitnii systemi universytetu]. *Cybersecurity: Education, Science, Technique, 3*(27), 72–86. https://doi.org/10.28925/2663-4023.2025.27.709
3. Lakhno, M. V. (2025). Contextual characteristics of digital footprints and their impact on university information security. *Central Ukrainian Scientific Bulletin. Technical Sciences, 11*(42, Pt. II), 11–22. https://doi.org/10.32515/2664-262X.2025.11(42).2.11-22
4. Lakhno, M. V. (2025). Method of multilevel analysis of digital footprints in information and educational systems [Metod bahatorivnevoho analizu tsyfrovykh slidiv v informatsiino-osvitnikh systemakh]. *Technical Sciences and Technologies, 3*(41), 193–202.
5. Shkarupylo, V. V., & Lakhno, M. V. (2025). Model of digital footprint analysis in secure information and educational systems. *Electronic Modeling, 47*(4), 113–125. https://doi.org/10.15407/emodel.47.04.113
6. Buitrago-Ropero, M. E., Ramírez-Montoya, M. S., & Laverde, A. C. (2023). Digital footprints (2005–2019): A systematic mapping of studies in education. *Interactive Learning Environments, 31*(2), 876–889. https://doi.org/10.1080/10494820.2020.1817509
7. Azcona, D., Hsiao, I. H., & Smeaton, A. F. (2019). Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints. *User Modeling and User-Adapted Interaction, 29*, 759–788. https://doi.org/10.1007/s11257-019-09227-7
8. Sun, L., Versteeg, S., Boztaş, S., & Rao, A. (2016). Detecting anomalous user behavior using an extended isolation forest algorithm: An enterprise case study. *arXiv*. https://arxiv.org/abs/1609.06676
9. Shi, L., Qian, C., & Guo, F. (2022). Real-time driving risk assessment using deep learning with XGBoost. *Accident Analysis & Prevention, 178*, 106836. https://doi.org/10.1016/j.aap.2022.106836
10. Folino, G., Otranto Godano, C., & Pisani, F. S. (2023). An ensemble-based framework for user behaviour anomaly detection and classification for cybersecurity. *The Journal of Supercomputing, 79*(11), 11660–11683. https://doi.org/10.1007/s11227-023-05230-9
11. Alaca, Y., Çelik, Y., & Goel, S. (2023). Anomaly detection in cyber security with graph-based LSTM in log analysis. *Chaos Theory and Applications, 5*(3), 188–197.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4 (32), 2026
ISSN 2663 – 4023

Myroslav, Lakhno
Postgraduate student,
Department of Computer Systems, Networks, and Cybersecurity,
National University of Life and Environmental Sciences of Ukraine, Kiev, Ukraine
ORCID: 0000-0001-6979-6076
Valss725@gmail.com

EXPERIMENTAL EVALUATION OF THE EFFECTIVENESS OF HYBRID METHODS FOR DIGITAL FOOTPRINT ANALYSIS IN DETECTING ATYPICAL BEHAVIOR IN INFORMATION AND EDUCATIONAL SYSTEMS

Abstract. The relevance of this study is driven by the need to strengthen the security of information and educational systems (IES) of higher education institutions (HEIs), which, under martial law and widespread distance learning, have also become targets of cyberattacks. Existing IES protection methods and tools based on static signatures and access control policies have lost their effectiveness against insider threats and behavioral anomalies, such as account compromise, academic misconduct, and unauthorized delegation of privileges. The aim of this work is an experimental evaluation of the effectiveness of the developed information technology for detecting atypical user activity through hybrid analysis of their digital footprints (DF). The study is based on the hypothesis that combining structural–hierarchical modeling of business processes with ensemble machine learning (ML) methods makes it possible to significantly reduce Type I and Type II errors. To verify the proposed solutions, a series of computational experiments was conducted on a real-world dataset formed from LMS Moodle log files (over 30,000 interaction events). A comparative analysis of the developed hybrid method with classical algorithms-XGBoost gradient boosting and Isolation Forest-was performed. Experimental results demonstrate that the proposed hybrid method, which employs weighted ensemble ($\omega_{xgb} = 0,95, \quad \omega_{iso} = 0,05$) learning, exhibits higher discriminative power and stability. The integral quality metric ROC-AUC reached 0,956, while the balanced F1-score achieved 0,858, exceeding the baseline XGBoost performance by 4,6%. Analysis of the Precision–Recall curves confirmed the robustness of the method to class imbalance, with the area under the curve (AP) equal to 0,889. The results of the study confirm that the implementation of the proposed technology enables higher education institutions to provide flexible protection of their information and educational systems by forming a clear separation between legitimate and atypical user behavior, while minimizing the risk of blocking bona fide users.

Keywords: digital footprints; information and educational system; anomaly detection; hybrid methods; machine learning; XGBoost; Isolation Forest; behavioral analysis; cybersecurity; higher education institution

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Dolliver, D. S., Ghazi-Tehrani, A. K., & Poorman, K. T. (2021). Building a robust cyberthreat profile for institutions of higher education: An empirical analysis of external cyberattacks against a large university's computer network. *International Journal of Law, Crime and Justice, 66*, 100484. https://doi.org/10.1016/j.ijlcj.2021.100484
2. Lakhno, M. (2025). System analysis of digital footprints in the information and educational system of a university [Systemnyi analiz tsyfrovykh slidiv u informatsiino-osvitnii systemi universytetu]. *Cybersecurity: Education, Science, Technique, 3*(27), 72–86. https://doi.org/10.28925/2663-4023.2025.27.709
3. Lakhno, M. V. (2025). Contextual characteristics of digital footprints and their impact on university information security. *Central Ukrainian Scientific Bulletin. Technical Sciences, 11*(42, Pt. II), 11–22. https://doi.org/10.32515/2664-262X.2025.11(42).2.11-22
4. Lakhno, M. V. (2025). Method of multilevel analysis of digital footprints in information and educational systems [Metod bahatorivnevoho analizu tsyfrovykh slidiv v informatsiino-osvitnikh systemakh]. *Technical Sciences and Technologies, 3*(41), 193–202.

5. Shkarupylo, V. V., & Lakhno, M. V. (2025). Model of digital footprint analysis in secure information and educational systems. *Electronic Modeling*, *47*(4), 113–125. https://doi.org/10.15407/emodel.47.04.113
6. Buitrago-Ropero, M. E., Ramírez-Montoya, M. S., & Laverde, A. C. (2023). Digital footprints (2005–2019): A systematic mapping of studies in education. *Interactive Learning Environments*, *31*(2), 876–889. https://doi.org/10.1080/10494820.2020.1817509
7. Azcona, D., Hsiao, I. H., & Smeaton, A. F. (2019). Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints. *User Modeling and User-Adapted Interaction*, *29*, 759–788. https://doi.org/10.1007/s11257-019-09227-7
8. Sun, L., Versteeg, S., Boztaş, S., & Rao, A. (2016). Detecting anomalous user behavior using an extended isolation forest algorithm: An enterprise case study. *arXiv*. https://arxiv.org/abs/1609.06676
9. Shi, L., Qian, C., & Guo, F. (2022). Real-time driving risk assessment using deep learning with XGBoost. *Accident Analysis & Prevention*, *178*, 106836. https://doi.org/10.1016/j.aap.2022.106836
10. Folino, G., Otranto Godano, C., & Pisani, F. S. (2023). An ensemble-based framework for user behaviour anomaly detection and classification for cybersecurity. *The Journal of Supercomputing*, *79*(11), 11660–11683. https://doi.org/10.1007/s11227-023-05230-9
11. Alaca, Y., Çelik, Y., & Goel, S. (2023). Anomaly detection in cyber security with graph-based LSTM in log analysis. *Chaos Theory and Applications*, *5*(3), 188–197.

Отримано редакцією журналу / Received: 04.01.26
Прорецензовано / Revised: 21.02.26
Схвалено до друку / Accepted: 26.03.26