DOI: https://doi.org/10.25140/2411-5363-2025-3(41)-193-202
УДК 004.056.55:378

***Мирослав Валерійович Лахно*[1]**

[1]аспірант кафедри комп'ютерних систем, мереж та кібербезпеки
Національний університет біоресурсів і природокористування України (Київ, Україна)
E-mail: valss725@gmail.com. **ORCID:** https://orcid.org/0000-0001-6979-6076. **ResearcherID:** KFO-8669-2024

МЕТОД БАГАТОРІВНЕВОГО АНАЛІЗУ ЦИФРОВИХ СЛІДІВ В ІНФОРМАЦІЙНО-ОСВІТНІХ СИСТЕМАХ

Запропоновано багаторівневий метод аналізу цифрових слідів користувачів інформаційно-освітніх систем з метою збільшення рівня інформаційної безпеки. Метод поєднує кластеризацію поведінкових профілів користувачів, баєсівське ймовірнісне моделювання та алгоритм машинного навчання XGBoost з акцентом на експоненціальне зважування якості кластеризації. Реалізована на Python система забезпечує виявлення ризиків та підтримку прийняття рішень щодо безпечного доступу до ресурсів інформаційно-освітніх систем закладів вищої освіти.

Ключові слова: *цифрові сліди; інформаційна безпека; інформаційно-освітня система; метод; кластеризація; XGBoost; баєсівська мережа; ризик.*

Рис.: 3. Бібл.: 12.

Актуальність теми дослідження. Інформаційно-освітні системи (ІОС) закладів вищої освіти (ЗВО) все частіше стають об'єктом кібернетичних атак через зростання обсягів особистих даних, цифрових ресурсів та інтенсивності взаємодії користувачів у віртуальному середовищі [1]. Ось чому головним завданням залишилося своєчасне виявлення загроз інформаційній безпеці (ІБ), які маскують під типову активність здобувачів вищої освіти (ЗдВО), викладачів та адміністративного персоналу ЗВО. Наявні методи аналізу поведінки користувачів [2] не беруть повністю до уваги гетерогенність цифрових слідів (ЦС) та контекстуальні фактори взаємодії з ІОС ЗВО. Це не дозволяє швидко адаптуватися до змін поведінкових шаблонів користувачів ІОС. Специфікою ІОС ЗВО є значна різноманітність ролей користувачів. Це ЗдВО, викладачі, адміністративний персонал, служби ЗВО тощо. Це зумовлює специфіку ЦС та підвищує ризики порушень ІБ. У більшості наявних робіт ЦС аналізуються переважно для освітньо-педагогічної постановки завдань. Зокрема, це ідентифікація стилів навчання, прогнозування успішності тощо. Тоді як проблематика інформаційних ризиків, пов'язаних із виявленням атипових поведінкових шаблонів, досі недостатньо опрацьована. Саме тому запропонований метод сфокусовано на оцінюванні рівня інформаційних ризиків, враховуючи специфіку ІОС ЗВО. У зв'язку з цим виникає нагальна потреба у створенні нових моделей та методів аналізу ЦС, здатних інтегрувати кластерний аналіз, ймовірнісне моделювання та прогнозування рівня ризиків ІБ у єдину систему підтримки прийняття рішень у сфері ІБ ЗВО.

Постановка проблеми. Гарантування інформаційної безпеки в ІОС ЗВО потребує нових підходів до спостереження за поведінкою користувачів на основі ЦС. Наявні рішення зазвичай реалізують окремі складники поведінкового аналізу. Також не враховують якість кластеризації, ймовірнісний характер взаємозв'язків між діями користувачів та подіями ІБ, а також мінливість рівня ризиків у часі. Такий підхід зменшує точність оцінки безпекових інцидентів і ускладнює адаптацію до нових сценаріїв загроз ІБ. Через це постає проблема розробки нового методу аналізу ЦС, який дозволить врахувати поведінкові відхилення, динаміку рівня ризиків та забезпечувати підтримку ухвалення рішень з ІБ ІОС.

Аналіз останніх досліджень і публікацій. Аналіз літературних джерел показав зростання зацікавленості до вивчення цифрових слідів (ЦС), зокрема у ЗВО. Так, M. Camacho у [3] дослідив роль ЦС у формуванні цифрової ідентичності здобувачів вищої освіти (ЗдВО). У [4] S. Karabatak проаналізував риси дій поведінки молоді у цифровому середовищі. У [5] M. Buitrago-Ropero систематизував дослідження щодо використання ЦС в освіті. H. Kumar у [6] дослідив рівень обізнаності ЗдВО щодо наслідків цифрової активності та виокремив проблему недостатньої підготовки до управління цифровою ідентичністю. У [7] A. Nugumanova дослідив математичні методи для виявлення

поведінкових шаблонів на основі ЦС на освітніх платформах та у соціальних мережах. У [8] C. Olipas оцінив рівень цифрової грамотності ЗдВО на основі аналізу ЦС. A. Songsom [9] запропонував модель управління освітніми відносинами на основі ЦС та Інтернету речей. У [10] M. Ollagnier-Beldame вивчав методи застосування ЦС для раннього виявлення ризику неуспішності на курсах програмування.

Виділення недосліджених частин загальної проблеми. Попри активне вивчення ЦС, більша частка розглянутих досліджень зосереджена на педагогічних аспектах, таких як формування цифрової ідентичності, прогнозування успішності або персоналізація навчання. При цьому недостатньо уваги приділено проблематиці інформаційної безпеки, зокрема виявленню загроз, що ґрунтується на комплексному аналізі поведінкових шаблонів дій користувачів інформаційно-освітніх платформ. Також відсутні узагальнені методи для багаторівневого аналізу цифрових слідів, які б поєднували кластеризацію, ймовірнісне моделювання та машинне навчання з урахуванням якісних характеристик опрацювання даних. До того ж не розроблено методів, що враховують зміни рівня ризику для оцінювання поведінкової мінливості користувачів у середовищі освітніх платформ. Залишилось відкритим питання інтеграції локальних моделей виявлення атипової поведінки із глобальною оцінкою рівня ризику для інформаційно-освітньої системи закладу вищої освіти.

Метою статті є розробка методу багаторівневого аналізу цифрових слідів в інформаційно-освітніх системах, який забезпечить виявлення атипової поведінки користувачів, оцінювання рівня інформаційних ризиків та підтримку ухвалення рішень щодо безпечного доступу до цифрових ресурсів закладу вищої освіти.

Виклад основного матеріалу. Основна ідея методу полягає в побудові ансамблю кластеризацій ЦС, результати якого зважуються за якістю за допомогою експоненціальної функції. Це дало змогу взяти до уваги гетерогенність поведінки користувачів у ІОС та підвищити релевантність групування даних. На базі кластеризованих підмножин ЦС формуємо локальні баєсівські мережі (БМ або BN). Ці баєсівські мережі відображають причинно-імовірнісні залежності між поведінковими параметрами. процесі дослідження це дозволило провести оцінку ймовірності виникнення інцидентів ІБ. Додатково використано алгоритм машинного навчання (ML) XGBoost для передбачення рівня ризику ІБ для ІОС. Інтегрована оцінка рівня ризику обчислена як зважене поєднання наслідків ймовірнісного моделювання та ML-прогнозування.

Формалізуємо метод. Нехай $D=\left\{x^{(1)},x^{(2)},...,x^{(N)}\right\}$, $x^{(i)}\in R^d$, множина ЦС, користувачів ІОС. Кожен вектор $x^{(i)}=\left(x_1^{(i)},x_2^{(i)},...,x_d^{(i)}\right)$ описує поведінкові характеристики сеансу користувача. Запроваджено ансамбль кластеризаторів $C=\left\{C^1,C^2,...,C^m\right\}$, кожен з яких формує кластеризацію $C^{(j)}:D\to\left\{1,...,K_j\right\}$. Якість кожного кластеризатора оцінюємо метрикою $Q^{(j)}>0.$ Ваги визначаємо експоненційно

$$\omega^{(j)}=\frac{e^{\beta\cdot Q^{(j)}}}{\sum_{t=1}^{m}e^{\beta\cdot Q^{(t)}}}, \tag{1}$$

де $\omega^{(j)}$ – вага j -го алгоритму кластеризації;

$\beta>0$ – гіперпараметр чутливості до різниці в якості кластерного аналізу (КА) $Q^{(j)}$. Тобто чим вища якість кластеризації $Q^{(j)}$, тим експоненційно більший її внесок у консенсус.

Вага $\omega^{(j)}$ кожного алгоритму кластеризації відображала його відносний внесок у підсумкове групування даних. Це означає, що алгоритм, який забезпечує вищу якість кластеризації за обраною метрикою, отримує експоненційно більший вплив на формування консенсусного рішення. У ролі метрики можна використати силуетний коефіцієнт або Davies-

Bouldin Index. Отже, ансамбль кластеризацій забезпечує гнучкий вибір алгоритмів і їх комбінацій. При цьому зберігає баланс між різними підходами до структурування даних.

Далі формуємо матрицю спільного членства:

$$M_{i,j} = \frac{1}{m}\sum_{t=1}^{m} \mathbb{I}\left[C^{(t)}\left(x^{(i)}\right) = C^{(t)}\left(x^{(j)}\right)\right], \tag{2}$$

де $\mathbb{I}[\cdot]$ – індикаторна функція.

Для кожного кластера $C_k \subset D$ будуємо БМ:

$$B_k = \left(G_k, \theta_k\right), \tag{3}$$

де $G_k = \left(V_k, E_k\right)$ – орієнтований ациклічний граф між змінними (ознаками) $x_1, ..., x_d$;

V_k – множина вузлів (поведінкових ознак користувачів);

E_k – множина дуг (залежностей);

$\theta_k = \{P(X_i | Pa(X_i))\}_{X_i \in V_i}$ – набір умовних ймовірнісних розподілів, визначених для кожної змінної X_i з множиною батьківських змінних $Pa\left(X_i\right) \subseteq V_k$.

Ймовірність ЦС у межах кластера визначено так:

$$P\left(x \middle| B_k\right) = \prod_{i=1}^{d} P\left(x_i \middle| Pa\left(x_i\right)\right). \tag{4}$$

Прогнозування рівня ризику здійснено за допомогою ML. Для всіх векторів $x \in D$ побудовано модель машинного навчання:

$$\hat{R}\left(x\right) = f_{ML}\left(x\right), \tag{5}$$

де f_{ML} – XGBoost ансамбль дерев.

Далі обчислюємо зважену оцінку рівня ризику ІБ ІОС з урахуванням БМ та ML-прогнозу:

$$R\left(x\right) = \lambda \cdot P\left(x \middle| B_k\right) + \left(1-\lambda\right) \cdot \hat{R}\left(x\right), \quad \lambda \in \left[0,1\right], \tag{6}$$

де $R\left(x\right) \in \left[0,1\right]$ – інтегрована оцінка рівня ризику дії користувача;

$\lambda \in \left[0,1\right]$ – ваговий коефіцієнт для збалансування обох компонент;

$P\left(x | B_k\right)$ – апостеріорна ймовірність поведінки згідно з кластерною баєсівською моделлю;

$\hat{R}\left(x\right)$ – прогнозоване значення рівня ризику за допомогою ML-моделі.

Отже, формулу (6) реалізовано у вигляді об’єднання прогнозу. Прогноз отримано за допомогою БМ та оцінки ризику з моделі XGBoost. Псевдокод, який ми наведемо далі (після рис. 1, див. рядки 13–16), деталізує цей процес.

На основі інтегрованої оцінки застосовуємо логіку доступу:

$$\delta\left(x_t\right) = \begin{cases} \text{"}дозволити\ доступ\text{"}, якщо\ R\left(x\right) < r_1; \\ \text{"}запит\ підтвердження\text{"}, якщо\ r_1 \le R\left(x\right) < r_2; \\ \text{"}заблокувати\ доступ\text{"}, якщо\ R\left(x\right) \ge r_2, \end{cases} \tag{7}$$

де r_1, r_2 – порогові значення для рівнів ризику (визначалися емпірично).

Узагальнену блок-схему роботи методу аналізу ЦС подано на рис. 1.

Через багаторівневу структуру та наявність декількох складників із незалежною логікою (кластеризація, БМ, XGBoost, контекстуальні профілі) формальне представлення методу у вигляді псевдокоду зафіксує логіку переходів між рівнями аналізу та відображенням результатів у кінцеву дію системи підтримки рішення.

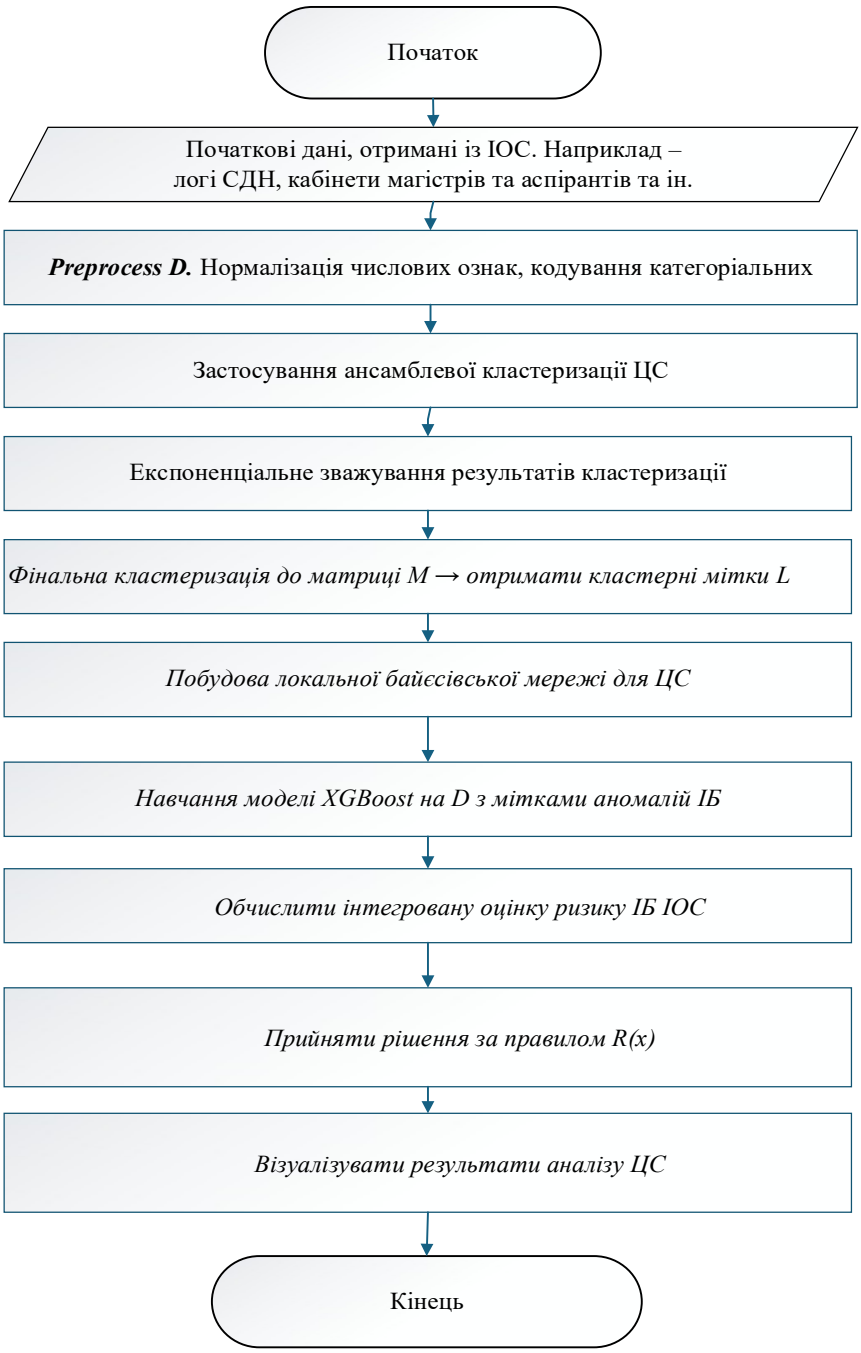

Рис. 1. Блок-схема методу аналізу ЦС

Джерело: розроблено автором.

Вхід:

X – множина ЦС користувачів ІОС ЗВО;
UP – профілі користувачів;
AR – база правил інцидентів;
T – часове вікно спостереження;

Вихід:
R – інтегрована оцінка рівня ризику користувача;*
D – множина прийнятих рішень (D ⊂ {Allow, Alert, Block});
01: ***Begin***
02: Perform preprocessing and normalization of digital traces in X;
03: Apply ensemble clustering to X to obtain clusters C = {C₁, C₂, ..., C_k};
04: ***For*** *each cluster C_i in C* ***do***
05: Construct a local Bayesian network B_i from samples in C_i;
06: ***For*** *each user x ∈ C_i* ***do***
07: Compute posterior probability P(x | B_i) using inference;
08: Evaluate risk score $R_\beta(x) = f(P(x \mid B_i))$;
09: ***End for***
10: ***End for***
11: Train XGBoost model F_XGB on labeled digital traces;
12: ***For*** *each user x ∈ X* ***do***
13: Predict refined risk score R_xgb(x) = F_XGB(z_x);
14: Obtain contextual profile CON(x) from UP;
15: Compute contextual risk adjustment $R\_ctx(x) = \alpha \cdot R_\beta(x) + (1-\alpha) \cdot \bar{R}(CON(x))$;
16: Integrate final risk $R^*(x) = \gamma \cdot R\_xgb(x) + (1-\gamma) \cdot R\_ctx(x)$;
17: ***If*** *R*(x) > threshold r_1* ***then***
18: Retrieve AR(x) from AR if exists matching rule;*
19: Make decision D(x) ∈ {Alert, Block} based on R and AR*(x);*
20: ***Else***
21: D(x) ← Allow;
22: ***End if***
23: ***End for***
24: ***End***

Псевдокод втілює багаторівневе опрацювання ЦС, яке починається з їх попередньої обробки шляхом нормалізації, фільтрації та векторизації ознак дій користувачів. Далі застосовується ансамбль кластеризації (K-Means, DBSCAN тощо). А результати агрегуються з урахуванням ваг, визначених за якістю кластеризації. Для кожного кластеру формуємо локальну баєсівську мережу (БМ). Вона дає змогу оцінити апостеріорну ймовірність рівні ризику ІБ для ІОС. Паралельно виконуємо навчання моделі XGBoost на розмічених даних. Далі надходять з навчальних платформ ЗВО або системи моніторингу подій ІБ. Отриманий прогноз уточнювався з урахуванням групової поведінки користувача. Фінальну інтегровану оцінку ризику формуємо як поєднання індивідуального прогнозу та контекстних характеристик дій користувача в ІОС. Рішення про подальші дії ухвалюємо на основі порогових значень рівня ризику та експертних правил, які зберігаються в базі знань системи підтримки прийняття рішень.

Урахування гетерогенності поведінки користувачів досягнуто завдяки використанню ансамблю кластеризацій, що включав алгоритми K-Means, DBSCAN та агломеративну кластеризацію. Вибір цих методів обґрунтований різною чутливістю до параметрів і здатністю виявляти як компактні (K-Means), так і довільної форми кластери (DBSCAN), а також ієрархічні структури (агломеративний підхід). Для оцінки якості кластеризації використовувалися метрики силуетного коефіцієнта та індексу Davies-Bouldin. Це дозволило встановити релевантність виділених груп. Метод аналізу ЦС реалізовано на мові програмування Python (v3.10).

Для експериментальних досліджень використовувалися дані, отримані з системи дистанційного навчання (Moodle) Національного університету біоресурсів і природокористування України. У табл. 1 висвітлено набір полів, кожне з яких відображає ознаки поведінкового профілю користувача на платформі https://elearn.nubip.edu.ua/. Числові атрибути відображали тривалість сесії (що моделює експоненціальний розподіл) та кількість виконаних дій (узгоджено з розподілом Пуассона). Це дозволило під час настових досліджень програмного продукту, створеного на основі методу аналізу ЦС, сформувати кількісну оцінку інтенсивності взаємодії з платформою Moodle.

Категоріальні характеристики містять тип пристрою, за рахунок якого відбувається взаємодія, час доби, а також роль користувача в освітньому середовищі. Ці дані сформували контекстуальне тло цифрової поведінки користувачів в ІОС ЗВО (на прикладі НУБІП України).

Таблиця 1 містить також бінарні індикатори – ознаки змін IP-адреси та географічної локації користувачів, або зміни контексту дій користувача на https://elearn.nubip.edu.ua/. Для тестового набору додано поле anomaly_label, яке відображало вручну або автоматично призначену анотацію. Це поле відображало, чи містила така сесія ознаки загрози ІБ (тобто рівень ризику) або відхилення від очікуваної поведінки. На рис. 1 представлено результат візуалізації кластерів ЦС користувачів ІОС https://elearn.nubip.edu.ua/, отриманих внаслідок застосування ансамблевої кластеризації та подальшого зменшення розмірності методом t-SNE. Кожна точка на площині (рис. 1) – це окремий сеанс користувача ІОС. Сеанс, відповідно, описано низкою поведінкових характеристик, попередньо нормалізованих та закодованих.

Різні кольори на рис. 1 вказували належність до відповідного кластеру. Кластер сформовано за допомогою поєднання результатів функціонування алгоритмів KMeans, DBSCAN і агломеративної кластеризації з використанням експоненційного зважування за метрикою якості. Просторове розташування кластерів на площині свідчить про наявність структурованих груп поведінки користувачів, які відрізнялися від типових ЦС. Видима відокремленість окремих скупчень вказує на ефективність методу для виявлення атипових шаблонів поведінки за допомогою аналізу ЦС.

Подібна структура тестового датасету дозволила інтегрувати часові, просторові, контекстуальні й поведінкові риси ЦС кожного користувача ресурсу Moodle.

Таблиця 1 – Структура тестового набору даних для дослідження моделі аналізу ЦС

Поле	Опис
session_duration	Тривалість сесії (у хвилинах. Для тесту взято з експоненційного розподілу)
num_actions	Кількість дій за сесію (пуассонівський розподіл)
device_type	Тип пристрою, за рахунок якого відбулася взаємодія (desktop / laptop / mobile / tablet)
time_of_day	Час доби (morning / afternoon / evening / night)
ip_change	Факт зміни IP (0/1)
geo_shift	Геолокаційне переміщення (0/1)
user_role	Роль користувача (student / teacher / admin)
anomaly_label	Ознака аномалії (0 = норма, 1 = загроза)

Результати дослідження наведено на рис. 2 та 3.

На рис. 3 представлено розподіл інтегрованої оцінки рівня ризику для сесій користувачів ІОС. Для кожної сесії поєднано прогноз ризику за моделлю XGBoost із апостеріорною ймовірністю з локальної БМ, отриманої після кластеризації ЦС. Розподіл дозволив візуалізувати частку сеансів з низьким, середнім та високим рівнем ризику. Отриманий результат підтвердив ефективність виявлення атипової поведінки користувачів.

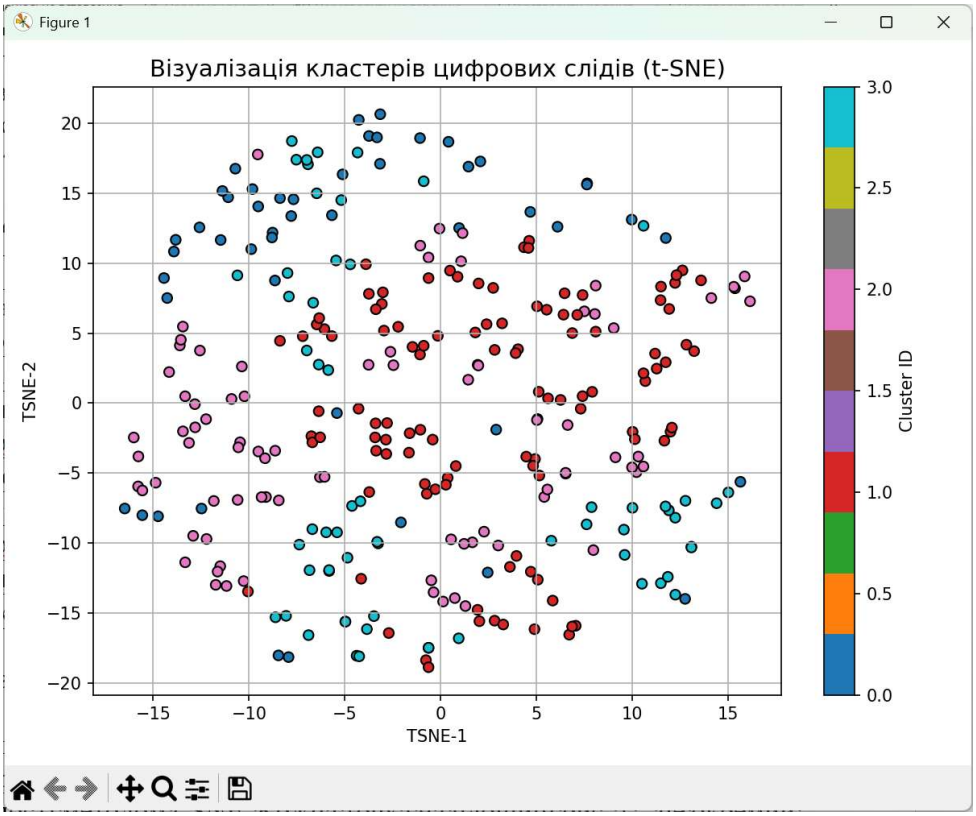

Рис. 2. Результат візуалізації кластерів ЦС користувачів ІОС

Як продемонстрував аналіз попередніх досліджень, більшість наявних підходів реалізують компоненти поведінкового аналізу окремо (табл. 2). Тобто не враховують взаємозв'язки між якістю кластеризації, достовірністю ймовірнісних висновків і продуктивністю прогнозних моделей дій користувачів ОІС.

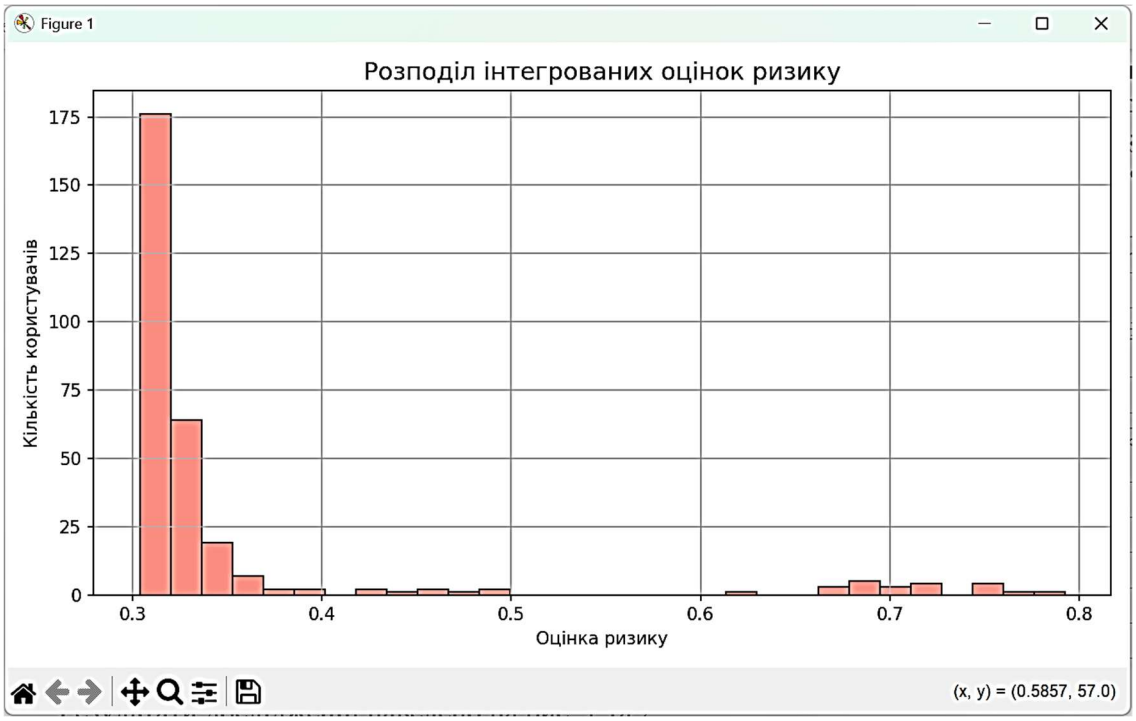

Рис. 3. Розподіл інтегрованої оцінки рівня ризику для сесій користувачів ІОС

Таблиця 2 – Функціональне порівняння методів та моделей аналізу ЦС

Критерій	Статичні ML-моделі	Моделі з фіксованою кластеризацією	Гібрид ML+BN (стандартний)	Новий метод
Урахування гетерогенності даних	частково	обмежене	частково	так
Адаптація до поведінкових змін	обмежена	неможлива	частково	реалізована
Якість кластеризації впливає на рішення	ні	ні	частково	так (через ваги)
Кластеризація як ансамбль	ні	ні	ні	так
Зважування за якістю кластеризації	–	–	–	Експоненційне
Інтерпретованість	низька	середня	висока	висока
Прогнозування ризику	так	ні	так	так
Комбіноване прийняття рішень	обмежено	ні	так	так

Джерело: систематизовано автором за даними [9-12].

Запропонований метод розширив межі методу гібридного аналізу ЦС завдяки введенню експоненційно зваженого ансамблю кластеризацій та урахування контекстно-залежної оцінки рівня ризику доступу користувачів ОІС на основі структурованого опрацювання поведінкових даних.

Висновки. У роботі вперше запропоновано новий метод багаторівневого аналізу цифрових слідів користувачів, що дає змогу підвищити ефективність виявлення атипової поведінки в інформаційно-освітньому середовищі закладу вищої освіти. Метод поєднує переваги кластеризації цифрових слідів, ймовірнісного моделювання та алгоритмів машинного навчання для формування системи оцінювання рівня ризиків інформаційної безпеки. Запропоновано застосувати ансамблеву кластеризацію з експоненційним зважуванням якості кластеризаторів. Результати експериментального дослідження засвідчили ефективність запропонованого методу. Візуалізація кластерної структури цифрових слідів підтвердила наявність поведінкових шаблонів дій користувачів, що корелюють з рівнем ризику інформаційної безпеки інформаційно-освітнього середовища закладу вищої освіти.

Наукова новизна роботи полягає у вперше запропонованому багаторівневому методі аналізу ЦС користувачів ІОС ЗВО, що відрізняється від відомих рішень поєднанням трьох складників: ансамблевої кластеризації з експоненціальним зважуванням якості алгоритмів; побудови локальних баєсівських мереж для відображення причинно-імовірнісних залежностей у поведінці користувачів; інтегрованого прогнозування рівня інформаційних ризиків на основі поєднання ймовірнісного моделювання та XGBoost. На відміну від наявних підходів, метод дозволив враховувати гетерогенність поведінки користувачів. Він забезпечив адаптацію до змін цифрових шаблонів у часі та підвищив точність виявлення атипових дій у середовищі ІОС.

Заява про використання генеративного ШІ та технологій на основі ШІ в процесі написання тексту статті.

Під час написання цього матеріалу автор використовував модель генеративного штучного інтелекту ChatGPT для узагальнення опису результатів експериментальних досліджень та оформлення списку використаних джерел. Після використання ChatGPT автор переглянув та відредагував зміст за потреби і взяв на себе повну відповідальність за зміст публікації.

Список використаних джерел

1. Ящук, І. (2021). Інформаційно-освітнє середовище закладу вищої освіти: практико-орієнтований підхід. *Інноватика у вихованні*, *1*(13), 62-72. https://doi.org/10.35619/iiu.v1i13.384.

2. Кіреєнко, О. Сценарії атак на систему дистанційної освіти. *Ukrainian Scientific Journal of Information Security*, *30*(2), 335-343. https://doi.org/10.18372/2225-5036.30.19246.

3. Camacho, M., Minelli, J., & Grosseck, G. (2012). Self and identity: raising undergraduate students' awareness on their digital footprints. *Procedia–Social and Behavioral Sciences*, *46*, 3176-3181. https://doi.org/10.1016/j.sbspro.2012.06.032.

4. Karabatak, S., & Karabatak, M. (2020). Z generation students and their digital footprints. *In 2020 8th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-5). https://doi.org/10.1109/ISDFS49300.2020.9116455.

5. 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/10.1080/10494820.2020.1814821.

6. Kumar, H., & Raj, P. (2020). An indagation on experiences and awareness of digital footprint among pupils of higher education. *Acad Res Int*, *11*(3), 16-31.

7. Nugumanova, A., & Baiburin, Y. (2019). Exploration of student behavior patterns through digital footprints. Journal of Mathematics, *Mechanics and Computer Science*, *103*(3), 43-54. https://doi.org/10.26577/JMMCS-2019-3-m5.

8. Olipas, C. N. P. (2023). The digital footprint awareness of the undergraduate students in a private higher education institution in nueva ecija, philippines: a basis for a plan of action. *International Advanced Research Journal in Science*, *Engineering and Technology*, 10 (2), 1528. https://doi.org/10.17148/IARJSET.2023.10203.

9. Songsom, N., Nilsook, P., Wannapiroon, P., Fung, C. C., & Wong, K. W. (2019). System design of a student relationship management system using the internet of things to collect the digital footprint. *International Journel of Emerging Technologies in Learning*, *14*(23), 125-140. https://doi.org/10.3991/ijet.v14i23.11066.

10. Ollagnier–Beldame, M. (2011). The use of digital traces: a promising basis for the design of adapted information systems. *International Journal on Computer Science and Information Systems*. *Special Issue "Users and Information Systems*", *6*(2), 24-45. https://hal.science/hal-00669019v1.

11. Whelan, E., Teigland, R., Vaast, E., & Butler, B. (2016). Expanding the horizons of digital social networks: Mixing big trace datasets with qualitative approaches. *Information and Organization*, *26*(1-2), 1-12. https://doi.org/10.1016/j.infoandorg.2016.03.001.

12. Sultan, M., Scholz, C., & van den Bos, W. (2023). Leaving traces behind: Using social media digital trace data to study adolescent wellbeing. *Computers in Human Behavior Reports*, *10*, 100281. https://doi.org/10.1016/j.chbr.2023.100281.

References

1. Yashchuk, I. (2021). Informatsiino-osvitnie seredovyshche zakladu vyshchoi osvity: prakyko-oriientovanyi pidkhid [Information and educational environment of higher education institution: practice-oriented approach]. *Innovatyka u vykhovanni – Innovations in education*, *1*(13), 62-72. https://doi.org/10.35619/iiu.v1i13.384.

2. Kirienko, O. (2021). Stsenarii atak na systemu dystantsiinoi osvity [Scenarios of attacks on the distance education system]. *Ukrainian Scientific Journal of Information Security*, *30*(2), 335-343. https://doi.org/10.18372/2225-5036.30.19246.

3. Camacho, M., Minelli, J., & Grosseck, G. (2012). Self and identity: Raising undergraduate students' awareness on their digital footprints. *Procedia–Social and Behavioral Sciences*, *46*, 3176-3181. https://doi.org/10.1016/j.sbspro.2012.06.032.

4. Karabatak, S., & Karabatak, M. (2020). Z generation students and their digital footprints. In 2020 *8th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-5). https://doi.org/10.1109/ISDFS49300.2020.9116455.

5. 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/10.1080/10494820.2020.1814821.

6. Kumar, H., & Raj, P. (2020). An indagation on experiences and awareness of digital footprint among pupils of higher education. *Academic Research International*, 11(3), 16-31. https://doi.org/10.26577/JMMCS-2019-3-m5.

7. Nugumanova, A., & Baiburin, Y. (2019). Exploration of student behavior patterns through digital footprints. *Journal of Mathematics, Mechanics and Computer Scienc*e, *103*(3), 43-54. https://doi.org/10.26577/JMMCS-2019-3-m5.

8. Olipas, C. N. P. (2023). The digital footprint awareness of the undergraduate students in a private higher education institution in Nueva Ecija, Philippines: A basis for a plan of action. International Advanced Research Journal in Science, *Engineering and Technology*, *10*(2), 15-28. https://doi.org/10.17148/IARJSET.2023.10203.

9. Songsom, N., Nilsook, P., Wannapiroon, P., Fung, C. C., & Wong, K. W. (2019). System design of a student relationship management system using the Internet of Things to collect the digital footprint. *International Journal of Emerging Technologies in Learning*, *14*(23), 125-140. https://doi.org/10.3991/ijet.v14i23.11066.

10. Ollagnier-Beldame, M. (2011). The use of digital traces: A promising basis for the design of adapted information systems. *International Journal on Computer Science and Information Systems. Special Issue "Users and Information Systems", 6*(2), 24-45. https://hal.science/hal-00669019v1.

11. Whelan, E., Teigland, R., Vaast, E., & Butler, B. (2016). Expanding the horizons of digital social networks: Mixing big trace datasets with qualitative approaches. *Information and Organization*, *26*(1-2), 1-12. https://doi.org/10.1016/j.infoandorg.2016.03.001.

12. Sultan, M., Scholz, C., & van den Bos, W. (2023). Leaving traces behind: Using social media digital trace data to study adolescent wellbeing. *Computers in Human Behavior Reports*, *10*, 100281. https://doi.org/10.1016/j.chbr.2023.100281

Отримано 06.07.2025

UDC 004.056.55:378

***Myroslav Lakhno*[1]**

[1]Postgraduate Student of the Department of Computer Systems
Networks and Cyder Security of the National University of Life and Environmental Sciences of Ukraine (Kyiv, Ukraine)
E-mail: valss725@gmail.com. **ORCID:** https://orcid.org/0000-0001-6979-6076. **ResearcherID:** KFO-8669-2024

A METHOD OF MULTILEVEL ANALYSIS OF DIGITAL TRACES IN INFORMATION AND EDUCATIONAL SYSTEMS

It is discussed in the article how to analyze digital footprints (DF) in the information and educational systems of higher education institutions (HEIs). A multilevel model of digital footprint analysis is proposed, which is aimed at improving the level of information security (IS) of a higher education institution by combining methods of clustering behavioral profiles, Bayesian networks, and machine learning algorithms, in particular XGBoost.

The peculiarity of the method is the use of an ensemble of clusters with exponential weighting of the results, which allows to consider the qualitative heterogeneity of clusters and to increase the accuracy of building user behavioral models in the information and educational systems of higher education institutions.

The method consists of several stages. The first stage involves preliminary processing of the CS. The second stage is ensemble clustering. At the third stage, local Bayesian networks are built for each cluster and the level of IS risk is predicted using XGBoost and the results are combined into a single integrated IS risk assessment for HEI. The decision on access to IOS resources is made based on risk thresholds. To verify the proposed method, an experimental model was implemented in Python 3.10 using the logs of the Moodle platform of the National University of Life and Environmental Sciences of Ukraine.

The proposed method provides for the detection of deviations in user's behavior and takes into account the heterogeneity of behavioral characteristics. The advantages of the model in comparison with existing methods are shown, in particular, by taking into account the quality of clustering, flexible interpretation of the risks level and integration of user's profiles into the decision-making process.

Keywords: *digital footprints; information security; information and education system; method; clustering; XGBoost; Bayesian network; risk.*

Fig.: 3. References: 12.

Лахно М. В. Метод багаторівневого аналізу цифрових слідів в інформаційно-освітніх системах. *Технічні науки та технології,* 2025. № 3(41). С. 193-202. DOI: https://doi.org/10.25140/2411-5363-2025-3(41)-193-202.