КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (23), 2024
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2024.23.3141
УДК 004.056

Лахно Валерій Анатолійович
доктор технічних наук, професор, професор кафедри комп'ютерних систем, мереж та кібербезпеки
Національний університет біоресурсів і природокористування України, Київ, Україна
ORCID 0000-0001-9695-4543
lva964@nubip.edu.ua

Волошин Семен Михайлович
кандидат технічних наук, доцент, доцент кафедри комп'ютерних систем, мереж та кібербезпеки
Національний університет біоресурсів і природокористування України, Київ, Україна
ORCID 0000-0002-4913-7003
voloshyn@nubip.edu.ua

Мамченко Сергій Миколайович
доктор педагогічних наук, професор, професор кафедри комп'ютерних систем, мереж та кібербезпеки
Національний університет біоресурсів і природокористування України, Київ, Україна
ORCID 0009-0006-8743-5606
s.mamchenko@nubip.edu.ua

Кулініч Олег Миколайович
кандидат технічних наук, доцент, доцент кафедри комп'ютерних систем, мереж та кібербезпеки
Національний університет біоресурсів і природокористування України, Київ, Україна
ORCID 0000-0002-0643-6898
o.kulinich@nubip.edu.ua

Касаткін Дмитро Юрійович
кандидат педагогічних наук, доцент, завідувач кафедри комп'ютерних систем, мереж та кібербезпеки
Національний університет біоресурсів і природокористування України, Київ, Україна
ORCID 0000-0002-2642-8908
d.kasatkin@nubip.edu.ua

КЛАСТЕРНИЙ АНАЛІЗ ДЛЯ ДОСЛІДЖЕННЯ ЦИФРОВИХ СЛІДІВ СТУДЕНТІВ ЗАКЛАДІВ ОСВІТИ

Анотація. Показано, що Кластерний Аналіз (КА) може використовуватися в процесі дослідження Цифрових Слідів (ЦС) студентів закладу освіти, а також інших закладів освіти, які впроваджують у навчальний процес Цифрове Освітнє Середовище (ЦОС). Кластерний аналіз може дозволити виявити патерни поведінки здобувачів освіти. Також застосування методів КА дозволить покращити персоналізацію навчання та підвищити ефективність освітніх програм. Показано, що в контексті забезпечення Інформаційної Безпеки (ІБ) ЦОС закладів освіти технології та методи аналізу ЦС також можуть бути корисними, наприклад, для: моніторингу мережевої активності студентів; аналізу журналів авторизації та автентифікації студентів; виявлення шкідливих програм та атак на ЦОС; аналізу загроз ІБ ЦОС в цілому; прогнозування вразливостей. Показано, що застосування методів КА може бути корисним при вивченні ступеня інформаційної безпеки ЦОС університетів та інших закладів освіти. Встановлено, що методи КА можуть допомогти виявляти групи студентів зі схожими образами активності з погляду ІБ, як ЦОС закладу освіти в цілому, так і його комп'ютерних мереж і систем. Встановлено, що з допомогою КА ЦС можна виявляти аномальну поведінку студентів, виявляти незвичайні патерни активності, факти несанкціонованого використання ресурсів чи інші відхилення від типової поведінки студентів у мережі закладу освіти. У статті також наведено результати експериментальних досліджень рівня компетентностей студентів різних спеціальностей в університеті з ІБ та захисту інформаційних активів ЦОС. У цьому були використані методи КА у процесі вивчення ЦС студентів. На основі КА ЦС різних груп студентів, зареєстрованих у ЦОС університету, було виділено шість типів користувачів. В результаті застосування методів КА

студенти, зареєстровані в ЦОС університету, були розбиті на відповідні кластери за критеріями, що впливають на ризики ІБ.

Ключові слова: цифрові сліди; кластерний аналіз; цифрове освітнє середовище закладу освіти; інформаційна безпека.

ВСТУП

Постановка проблеми. Останнім часом більшість закладів освіти у всьому світі, особливо на тлі стрімкого поширення пандемії Covid-19, крім традиційних форм навчання, пропонують студентам різні форми онлайн навчання. Такий формат навчання, наприклад, реалізують через Системи Дистанційного Навчання (СДН) чи онлайн курси з певних дисциплін. Однак для ефективного функціонування подібних онлайн курсів та СДН в цілому, необхідне відповідне цифрове середовище або Інформаційна Система (ІС), яке виконуватиме функції з управління та організації навчального процесу. У більшості випадків подібні завдання вирішуються за допомогою систем управління навчанням, а саме LMS, які пропонують безліч функцій для підтримки викладачів у процесі створення, адміністрування та управління онлайн курсами.

Однак такі системи у більшості випадків мають досить обмежений арсенал модулів для аналізу даних, що накопичуються в ході організації навчального процесу [1], [2]. Зокрема, у таких системах, недостатньо реалізовано потенціал для аналізу про Цифрових Слідів (ЦС) студентів під час навчання [3] – [5]. Подібні LMS системи практично не використовують накопичені дані щодо ЦС студентів для підвищення ступеня захищеності та інформаційної безпеки університету від зовнішніх втручань.

Зауважимо, що й у контексті забезпечення Інформаційної Безпеки (ІБ) ЦОС закладу освіти технології та методи Інтелектуального Аналізу Даних (ІАД) та ЦС також можуть бути корисними, наприклад, для: моніторингу мережевої активності студентів та викладачів; аналізу журналів авторизації та автентифікації студентів та викладачів; виявлення шкідливих програм та атак на ЦОС закладу освіти; аналізу загроз ІБ ЦОС закладу освіти загалом; прогнозування вразливостей.

Таким чином, все вище описано і визначає актуальність даного дослідження, а також мотивує нас до продовження роботи у даному напрямі.

Аналіз останніх досліджень і публікацій. В [4] авторами показано, що дослідження процесів навчання може виграти від вивчення ЦС, які залишають студенти, зокрема, при перегляді навчального контенту освітніх платформ. В [5] автори представляють результати досліджень з інтеграції стилів навчання на підставі аналізу ЦС в освітні вебсистеми на базі LMS Claroline, Ganesha, Chamilo, Moodle. Така інтеграція результатів аналізу ЦС, на думку авторів, допоможе студентам в засвоєнні навчального контенту курсу. Проте автори обмежилися лише порівняльним аналізом LMS систем, не деталізуючи методики застосування конкретного методу аналізу ЦС студентів у LMS системі.

В [6] автори, використавши методи кластерного аналізу даних, отриманих у тому числі цифровими слідами студентів, спробували вирішити завдання, пов’язане з шляхами підвищення ефективності освітнього процесу. Проте робота не розглядає багато технологічних аспектів аналізу ЦС студентів.

Використовуючи методи кореляційного та кластерного аналізу автори дослідження у [7] проаналізували стратегії освітньої діяльності студентів у різних соціальних мережах.

У працях [8] – [10] автори детально досліджували за допомогою різних методів ІАД залежності між успіхами студентів та їх активністю в онлайн курсах на базі LMS Moodle. Такі дослідження дозволили виявити поведінкові стратегії студентів під час онлайн навчання.

Динаміка зростання інцидентів ІБ у закладах освіти, які активно впроваджують в освітній процес інформаційні системи та технології, свідчить про еволюціонування такого роду загроз.

В дослідженнях [11] – [15] авторами розглядаються різні аспекти проблематики застосування методів ІАД для забезпечення ІБ ЦОС закладу освіти.

У роботах [16] – [18] автори досліджують проблематику зв'язку — «Цифровий слід» — «Інтелектуальний аналіз даних» — «Інформаційна безпека».

Як показав аналіз попередніх досліджень, Кластерний Аналіз (КА) може бути корисним інструментом для поділу студентів, зареєстрованих у ЦОС закладу освіти та, зокрема, в СДН, за рівнем їх технічних знань про ІБ, ризики та заходи безпечної роботи в ЦОС закладу освіти. Методи кластерного аналізу дозволять виділити групи студентів на основі подібних характеристик та потреб, що також може бути використане для формування індивідуальних освітніх програм.

Метою статті є виявити групи в ЦОС закладу освіти з найменшими контекстно-залежними характеристиками щодо питань дотримання правил ІБ при роботі в даному середовищі та рівнів ризиків ІБ при роботі різних категорій користувачів в ЦОС закладу освіти.

У зв'язку з цим необхідно вирішити наступні завдання:

1. ідентифікувати масив подібних груп за певними критеріями;
2. провести аналіз ЦС студентів, віднесених до цієї групи та провести кластерний аналіз на основі методу k-means.

МЕТОДИКА ДОСЛІДЖЕННЯ

Для виявлення груп студентів та викладачів у ЦОС закладу освіти з найменшими знаннями та досвідом у питаннях ІБ використовується аналіз ЦС та КА. Були послідовно реалізовані такі етапи:

Етап 1. Збір даних. Були зібрані доступні дані про користувачів, зареєстрованих у ЦОС університету з використанням відповідних логів LMS Moodle та ОС. Ці дані, описували, наприклад, проходження відповідних курсів з ІБ, ступінь та успішність (для студентів) виконання завдань з ІБ, результати тестувань тощо за відповідними курсами, що пов'язані з ІБ. Також ці дані отримані на підставі: оцінювання результатів тестування студентів за темами ІБ; моніторингу активності користувачів у мережі; оцінювання заходів захисту в ЦОС університету (наприклад, того, наскільки активно студенти використовують такі заходи захисту, як стійкі паролі, двофакторна автентифікація та шифрування даних, що свідчить про інформованість користувачів щодо загроз ІБ).

Етап 2. Підготовка даних. Зібрані дані були відфільтровані з метою видалення неповних записів. А потім перетворені на числовий формат.

Етап 3. Аналіз цифрових слідів: ЦС користувачів ЦОС університету можуть вказувати на ризики ІБ. До таких ЦС можна віднести: поведінкові дані (незвичайні спроби доступу, наприклад, до СДН, часті невдалі спроби входу в обліковий запис, несподівані зміни місця розташування або пристроїв для доступу до навчальних матеріалів; аномальна активність (незвичайно інтенсивне скачування або копіювання матеріалів, що виходить за рамки

звичного використання); відхилення від звичного розкладу (несподівані або незвичні періоди активності, входи в систему в незвичний час для конкретного студента); зміни у звичайних патернах поведінки; незвичайні запити на зміну облікових даних попереднього повідомлення або наявності підтвердження). Використовуючи методи аналізу ЦС, також було вилучено додаткову інформацію про пізнання, а також досвід студентів та викладачів з питань ІБ, наприклад, це може бути аналіз активності в ЦОС університету, стилю поведінки, участі у дискусійних форумах тощо.

Етап 4. Реалізація КА. Застосування методів КА для групування студентів на основі їх знань та досвіду з питань ІБ.

Етап 5. Інтерпретація результатів.

Зауважимо, що результати аналізу ЦС та КА можуть бути наближеними та вимагати додаткової інтерпретації, а також перевірки. Однак ці методи можуть допомогти керівникам різних рівнів закладу освіти розпочати процес виявлення груп студентів, які мають найменший рівень знань та досвіду з питань ІБ з метою підвищення ступеня захищеності ЦОС університету в цілому.

Позначимо через Ω всю множину користувачів, зареєстрованих у ЦОС університету. Для класифікації та кластеризації потрібно сформувати однорідні за ознакою кластери користувачів за рівнем ризиків для ІБ ЦОС університету. Для цього скористаємося методами кластерного аналізу — ієрархічною класифікацією та методом k-means.

У найпростішому варіанті в якості метрика можна використати евклідову відстань

$$\rho_{ik} = \sqrt{\sum_{i=1}^{m} \left(z_{ji} - z_{jk}\right)^2}, \qquad (1)$$

де ρ_{ik} — відстань між, відповідно, i,k спостереженнями, яка сформована на основі стандартизованих даних, що згадані вище; z_{ji} – матриця стандартизированих даних.

Виконаємо кластеризацію, використовуючи максимально наближені до дійсності дані групи користувачів, зареєстрованих в ЦОС університету.

Нехай множина Ω шість груп, що наведено у табл. 1. Дані отримано на основі аналізу ЦС користувачів, а саме студентів різних спеціальностей Національного університету біоресурсів та природокористування України (Україна, м. Київ).

Таблиця 1

Класифікація користувачів, зареєстрованих у ЦОС університету за критерієм ризиків для ІБ

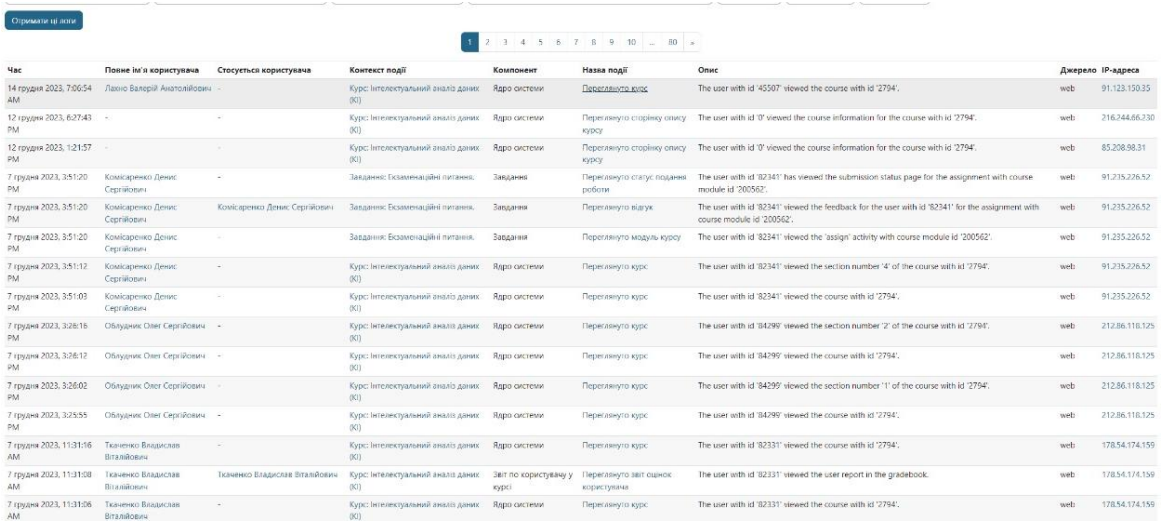
№	Назва групи (типи користувачів)	Ознаки (Опис моделі поведінки)
1	Обізнані користувачі	До цієї групи можна віднести користувачів, які добре знають ризики в мережі закладу освіти (або ЦОС закладу освіти в цілому). Ці користувачі вживають заходи для забезпечення ІБ своїх даних та облікових записів в ЦОС закладу освіти. Вони завжди дотримуються рекомендацій щодо створення складних паролів, регулярно оновлюють програмне забезпечення, не відкривають підозрілі покликання або вкладення в електронних листах, а також використовують надійне антивірусне Програмне Забезпечення (ПЗ) на своїх пристроях.
2	Недбалі користувачі	До групи віднесено користувачів, які не звертають належної уваги на заходи ІБ. Такі користувачі вразливі до атак. Користувачів цієї групи характеризує використання слабких паролів, повторення паролів для різних акаунтів, не своєчасне оновлення ПЗ, ігнорування підозрілої активності та заходів для захисту своїх даних у ЦОС закладу освіти.

ISSN 2663 - 4023

3	Незнаючі користувачі	Група містить користувачів, які не мають достатнього рівня знань про ІБ, ризики при роботі в мережі закладу освіти. Ця група може використовувати ненадійне ПЗ на своїх пристроях при підключенні до ЦОС закладу освіти, а також можуть перевати конфіденційні дані через незахищені канали зв'язку.
4	Байдужі користувачі	До цієї групи входять користувачі, які не виявляють інтересу до питань ІБ ЦОС закладу освіти. Ця категорія користувачів не перевіряє свої акаунти щодо злому, не звертає увагу на попередження про можливі загрози тощо.
5	Безвідповідальні користувачі	Учасники освітнього процесу порушують правила та політику ІБ у мережі закладу освіти. Вони можуть спробувати отримати несанкціонований доступ до систем ЦОС закладу освіти, поширювати шкідливе ПЗ, порушувати конфіденційність даних або вести недобросовісну активність у ЦОС закладу освіти.
6	Деструктивні користувачі	Ці користувачі намагаються завдати шкоди мережі закладу освіти, в тому числі поширюючи віруси, блокуючи ресурси мережі тощо.
Примітка. Ці групи користувачів досить умовні. Часто немає чітких меж між групами користувачів. Також можуть бути відмінності між групами. У міру набуття знань, наприклад, за рахунок відповідних курсів, користувачі можуть переходити від одного типу до іншого, усвідомлюючи важливість ІБ у мережі та вживаючи відповідних заходів для захисту своїх даних та облікових записів.		

Будь-яку групу користувачів можна охарактеризувати, використовуючи такі ознаки, приклад наведено рис. 1 та рис. 2:

1) Назва спеціальності та ID групи користувачів (NS_IDGroup — доступна, наприклад, з LMS Claroline, Ganesha, Chamilo, Moodle);
2) Курс/Рік навчання 1–4/5 (Year_ study — наприклад, 1 (1-й курс) і т.д.) (береться, наприклад, з LMS Claroline, Ganesha, Chamilo, Moodle));
3) Особиста відповідальність (Pers_resp (1–100)) — характеризується наявністю специфічних ознак, наприклад, ступінь акуратності в обробці особистих даних, частота зміни паролів, базові знання про ІБ, активність використання захищених з'єднань при роботі в ЦОС закладу тощо, тобто ці дані частково можна одержати на основі аналізу ЦС користувачів за допомогою LMS Claroline, Ganesha, Chamilo, Moodle, а також використовуючи SIEM);
4) Середня успішність за курсами, пов'язаними з ІТ та ІБ (Aver_ perf (0–100) — доступна, наприклад, з LMS Claroline, Ganesha, Chamilo, Moodle));
5) Оцінка компетентностей з ІБ (Comp_ ass (0-10), наприклад, наявність знань про: фішинг; шкідливе ПЗ; використання слабких/сильних паролів; оновлення ПЗ, включаючи операційну систему; безпеки мережі, в тому числі використання захищеного Wi-Fi; налаштування брандмауера та використання VPN; безпечне зберігання даних тощо, наприклад, береться на основі тестування та/або анкетування користувачів ЦОС закладу освіти).



Час	Повне ім'я користувача	Стосується користувача	Контекст події	Компонент	Назва події	Опис	Джерело	IP-адреса
14 грудня 2023, 7:00:54 AM	Лахно Валерій Анатолійович	-	Курс: Інтелектуальний аналіз даних (КІ)	Ядро системи	Переглянуто курс	The user with id '45507' viewed the course with id '2794'.	web	91.123.150.35
12 грудня 2023, 6:27:43 PM	-	-	Курс: Інтелектуальний аналіз даних (КІ)	Ядро системи	Переглянуто сторінку опису курсу	The user with id '0' viewed the course information for the course with id '2794'.	web	216.244.66.230
12 грудня 2023, 1:21:57 PM	-	-	Курс: Інтелектуальний аналіз даних (КІ)	Ядро системи	Переглянуто сторінку опису курсу	The user with id '0' viewed the course information for the course with id '2794'.	web	85.208.98.31
7 грудня 2023, 3:51:20 PM	Комісаренко Денис Сергійович	-	Завдання: Екзаменаційні питання.	Завдання	Переглянуто статус подання роботи	The user with id '82341' has viewed the submission status page for the assignment with course module id '200562'.	web	91.235.226.52
7 грудня 2023, 3:51:20 PM	Комісаренко Денис Сергійович	Комісаренко Денис Сергійович	Завдання: Екзаменаційні питання.	Завдання	Переглянуто відгук	The user with id '82341' viewed the feedback for the user with id '82341' for the assignment with course module id '200562'.	web	91.235.226.52
7 грудня 2023, 3:51:20 PM	Комісаренко Денис Сергійович	-	Завдання: Екзаменаційні питання.	Завдання	Переглянуто модуль курсу	The user with id '82341' viewed the 'assign' activity with course module id '200562'.	web	91.235.226.52
7 грудня 2023, 3:51:12 PM	Комісаренко Денис Сергійович	-	Курс: Інтелектуальний аналіз даних (КІ)	Ядро системи	Переглянуто курс	The user with id '82341' viewed the section number '4' of the course with id '2794'.	web	91.235.226.52
7 грудня 2023, 3:51:03 PM	Комісаренко Денис Сергійович	-	Курс: Інтелектуальний аналіз даних (КІ)	Ядро системи	Переглянуто курс	The user with id '82341' viewed the course with id '2794'.	web	91.235.226.52
7 грудня 2023, 3:26:16 PM	Облудник Олег Сергійович	-	Курс: Інтелектуальний аналіз даних (КІ)	Ядро системи	Переглянуто курс	The user with id '84299' viewed the section number '2' of the course with id '2794'.	web	212.86.118.125
7 грудня 2023, 3:26:12 PM	Облудник Олег Сергійович	-	Курс: Інтелектуальний аналіз даних (КІ)	Ядро системи	Переглянуто курс	The user with id '84299' viewed the course with id '2794'.	web	212.86.118.125
7 грудня 2023, 3:26:02 PM	Облудник Олег Сергійович	-	Курс: Інтелектуальний аналіз даних (КІ)	Ядро системи	Переглянуто курс	The user with id '84299' viewed the section number '1' of the course with id '2794'.	web	212.86.118.125
7 грудня 2023, 3:25:55 PM	Облудник Олег Сергійович	-	Курс: Інтелектуальний аналіз даних (КІ)	Ядро системи	Переглянуто курс	The user with id '84299' viewed the course with id '2794'.	web	212.86.118.125
7 грудня 2023, 11:31:16 AM	Ткаченко Владислав Віталійович	-	Курс: Інтелектуальний аналіз даних (КІ)	Ядро системи	Переглянуто курс	The user with id '82331' viewed the course with id '2794'.	web	178.54.174.159
7 грудня 2023, 11:31:08 AM	Ткаченко Владислав Віталійович	Ткаченко Владислав Віталійович	Курс: Інтелектуальний аналіз даних (КІ)	Звіт по користувачу у курсі	Переглянуто звіт оцінок користувача	The user with id '82331' viewed the user report in the gradebook.	web	178.54.174.159
7 грудня 2023, 11:31:06 AM	Ткаченко Владислав Віталійович	-	Курс: Інтелектуальний аналіз даних (КІ)	Ядро системи	Переглянуто курс	The user with id '82331' viewed the course with id '2794'.	web	178.54.174.159

Рис. 1. Скріншот логів, що містить інформацію про ЦС студентів (LMS Moodle, НУБіП України)

NS_IDGroup	Year_ study	Pers_ resp	Aver_ perf	Comp_ ass
Software Engineering_1501	1	21	68	4
Software Engineering_1501	2	40	74	5
Software Engineering_1501	3	56	77	6
Software Engineering_1501	4	65	84	6
Information Technology_1502	1	17	65	4
Information Technology_1502	2	37	77	5
Information Technology_1502	3	61	81	6
Information Technology_1502	4	71	96	6
Economic cybernetics_1503	1	21	71	4
Economic cybernetics_1503	2	44	76	5
Economic cybernetics_1503	3	60	80	6
Economic cybernetics_1503	4	71	74	7
Computer engineering_1504	1	22	69	4
Computer engineering_1504	2	42	75	5
Computer engineering_1504	3	61	74	7
Computer engineering_1504	4	75	78	9
Economy_1601	1	14	61	2
Economy_1601	2	29	73	3
Economy_1601	3	45	74	4
Economy_1601	4	50	75	4
Enterprise economy_1602	1	12	62	2
Enterprise economy_1602	2	25	73	3
Enterprise economy_1602	3	40	73	4
Enterprise economy_1602	4	49	75	4
Management_1701	1	11	65	2
Management_1701	2	27	72	3
Management_1701	3	39	73	3
Management_1701	4	42	75	4
Ecology_1604	1	12	66	2
Ecology_1604	2	26	73	3
Ecology_1604	3	34	72	3
Ecology_1604	4	41	80	4

NS_IDGroup	Year_ study	Pers_ resp	Aver_ perf	Comp_ ass
Software Engineering_1501	-1,32051	-0,97789	-0,87045	-0,20733
Software Engineering_1501	-0,44017	0,033262	0,032936	0,395806
Software Engineering_1501	0,44017	0,884758	0,484629	0,998939
Software Engineering_1501	1,320511	1,363725	1,538579	0,998939
Information Technology_1502	-1,32051	-1,19077	-1,32214	-0,20733
Information Technology_1502	-0,44017	-0,12639	0,484629	0,395806
Information Technology_1502	0,44017	1,150851	1,086886	0,998939
Information Technology_1502	1,320511	1,683037	3,345352	0,998939
Economic cybernetics_1503	-1,32051	-0,97789	-0,41876	-0,20733
Economic cybernetics_1503	-0,44017	0,246136	0,334065	0,395806
Economic cybernetics_1503	0,44017	1,097633	0,936322	0,998939
Economic cybernetics_1503	1,320511	1,683037	0,032936	1,602073
Computer engineering_1504	-1,32051	-0,92467	-0,71989	-0,20733
Computer engineering_1504	-0,44017	0,139699	0,1835	0,395806
Computer engineering_1504	0,44017	1,150851	0,032936	1,602073
Computer engineering_1504	1,320511	1,895911	0,635193	2,808339
Economy_1601	-1,32051	-1,35042	-1,9244	-1,41359
Economy_1601	-0,44017	-0,55214	-0,11763	-0,81046
Economy_1601	0,44017	0,299354	0,032936	-0,20733
Economy_1601	1,320511	0,565447	0,1835	-0,20733
Enterprise economy_1602	-1,32051	-1,45686	-1,77384	-1,41359
Enterprise economy_1602	-0,44017	-0,76502	-0,11763	-0,81046
Enterprise economy_1602	0,44017	0,033262	-0,11763	-0,20733
Enterprise economy_1602	1,320511	0,512229	0,1835	-0,20733
Management_1701	-1,32051	-1,51008	-1,32214	-1,41359
Management_1701	-0,44017	-0,65858	-0,26819	-0,81046
Management_1701	0,44017	-0,01996	-0,11763	-0,81046
Management_1701	1,320511	0,139699	0,1835	-0,20733
Ecology_1604	-1,32051	-1,45686	-1,17158	-1,41359
Ecology_1604	-0,44017	-0,7118	-0,11763	-0,81046
Ecology_1604	0,44017	-0,28605	-0,26819	-0,81046
Ecology_1604	1,320511	0,08648	0,936322	-0,20733

Рис. 2. Скріншоти таблиці в STATISTICA 12.5 з вихідними даними для аналізу та зі стандартизованими змінними

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Мета КА — розбити користувачів ЦОС закладу освіти на класи. Причому кожен із таких класів відповідає своїй групі ризику у контексті ІБ. Спостереження, що потрапили до однієї групи, характеризуються однаковою ймовірністю інциденту ІБ.

Дослідження проводилося з використанням пакету STATISTICA 12.5.

На першому етапі на основі ієрархічної класифікації були отримані дендрограми, необхідні для групування об'єктів у підмножини (кластери), на основі їх подібності (рис. 3).

ISSN 2663 - 4023

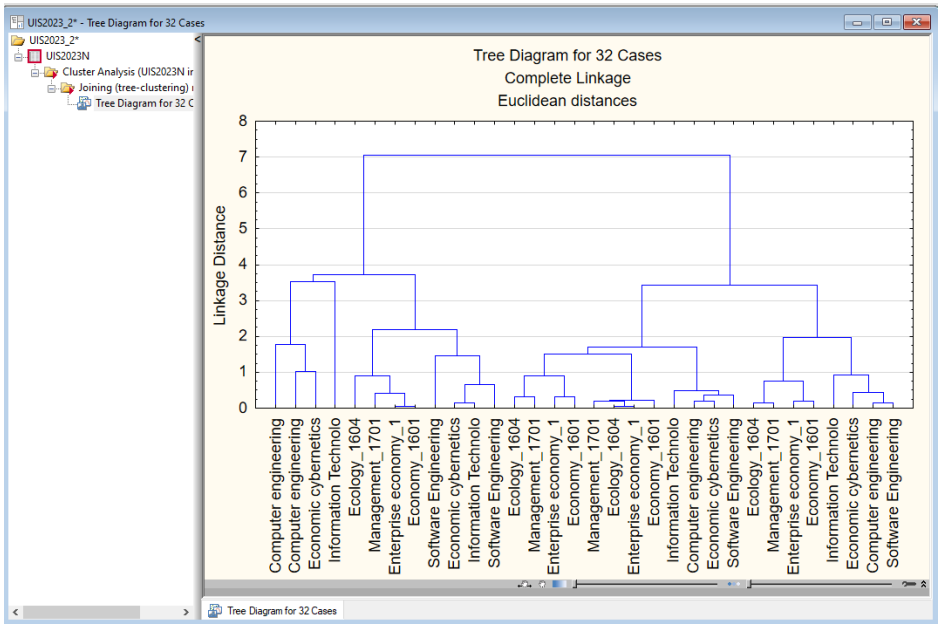


Рис. 3. Скріншот результатів ієрархічної класифікації (пакет STATISTICA 12.5)

Необхідно серед наявної множини груп виділити відповідні кластери. Наприклад, користувачі, віднесені до небезпечних з погляду критеріїв ІБ та ризиків для ЦОС закладу освіти, байдужі користувачі та нейтральні за оцінюваними ознаками.

Припущення. Дані про всі вище зазначені ознаки доступні і можуть бути оцінені та виміряні. Наприклад, зміна пароля, IP та інші ЦС можна відстежувати в системах LMS, а також за допомогою вже згаданих SIEM систем.

В результаті дослідження на основі методу k-means були отримані графіки середніх та довірчих інтервалів для змінних у кожному кластері, що характеризує користувачів ЦОС закладу освіти в контексті дотримання правил ІБ та впливу їхнього стилю роботи на ризики для ІБ (рис. 4).

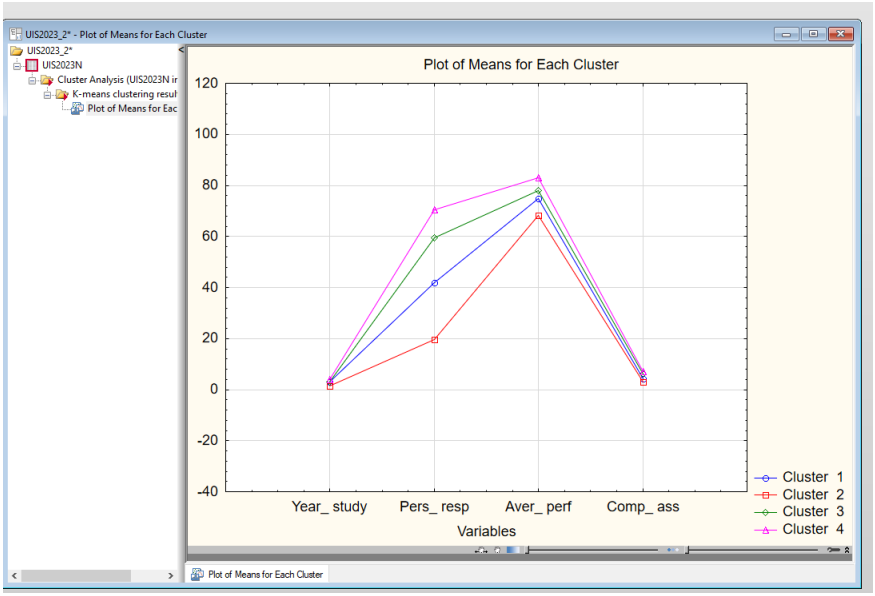


Рис. 4. Результати кластерного аналізу

Порівнявши метод k-means та ієрархічну класифікацію, необхідно зазначити, що безперечною перевагою першого методу є можливість роботи з первинними даними. Це дозволить, наприклад, фахівцям з ІБ закладів освіти, обробляти досить великі обсяги даних, які можна імпортувати з LMS у форматі *.csv.

У випадку значного інформаційного масиву користувачів великих закладів освіти це безсумнівна перевага. Більш того, метод k-means може компенсувати наслідки неякісного вихідного розбиття вихідного масиву даних.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У роботі наведено результати дослідження рівня компетентностей користувачів різних спеціальностей Цифрового Освітнього Середовища Університету (ЦОСУ) з питань, пов'язаних із ІБ. Дослідження проведено для Національного університету біоресурсів та природокористування України. Використовуються методи кластерного аналізу та аналізу ЦС користувачів ЦОС університету. На основі аналізу ЦС різних груп зареєстрованих користувачів у ЦОС університету було досліджено поведінку шести типів користувачів. Ці типи користувачів різних спеціальностей, що розглядаються в тестовому наборі, в результаті застосування ієрархічної класифікації та методу k-means були розбиті на відповідні кластери за критеріями, що впливають ризики ІБ ЦОС закладу освіти. Для кожного кластера експертом з ІБ ЦОС університету може визначатися ймовірність настання випадків, пов'язаних з високим рівнем ризику ІБ, та відповідно, можуть бути вжиті заходи для усунення причин таких випадків та розроблятись рекомендації для користувачів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Oliveira, P., Cunha, C., & Nakayama, M. (2016). Learning Management Systems (LMS) and E-learning Management: An Integrative Review and Research Agenda. *JISTEM-J. Inf. Syst. Technol. Manag. 13,* 157–180.
2. Aldiab, A., et al. (2019). Utilization of Learning Management Systems (LMSs) in Higher Education System: A Case Review for Saudi Arabia. *Energy Procedia, 160,* 731–737. https://doi.org/10.1016/j.egypro.2019.02.186
3. Azcona, D., Hsiao, I., & Smeaton, A. (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/s11257-019-09234-7
4. Nai, R., et al. (2023). Process Mining on Students' Web Learning Traces: A Case Study with an Ethnographic Analysis. *European Conference on Technology Enhanced Learning, Lecture Note in Computer Science, 14200,* 599–604. https://doi.org/10.1007/978-3-031-42682-7_48
5. Mohssine, B., et al. (2021). Adaptive Help System Based on Learners 'Digital Traces' and Learning Styles. *Int. J. Emerging Technol. Learning (iJET), 16(10),* 288–294. https://doi.org/10.3991/ijet.v16i10.19839
6. Ye, D., & Pennisi, S. (2022). Using Trace Data to Enhance Students' Self-Regulation: A Learning Analytics Perspective. *The Internet and Higher Education, 54, 100855.* https://doi.org/10.1016/j.iheduc.2022.100855
7. Noskova, T., Pavlova, T., & Yakovleva, O. (2018). Study of Students' Educational Activity Strategies in the Social Media Environment. *E-learning and Smart Learning Environment for the Preparation of New Generation Specialists, 10,* 113–125.
8. Kadoić, N., & Oreški, D. (2018). Analysis of Student Behavior and Success Based on Logs in Moodle. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO).* https://doi.org/10.23919/MIPRO.2018.8400123
9. Mogus, A., Djurdjevic, I., & Suvak, N. (2012). The Impact of Student Activity in a Virtual Learning Environment on Their Final Mark. *Active Learning in Higher Education, 13(3),* 177–189. https://doi.org/10.1177/1469787412452985
10. Stiller, K., & Bachmaier, R. (2018). Identifying Learner Types in Distance Training by Using Study Times. *EDEN Conference Proceedings, 1,* 78–86.

11. Ahmed, A., Alharthe, R., & Alfereej, M. (2023). Organizational Committees and Their Role in Enhancing Intellectual Security: A Case Study on Female Students of the Bachelor of Information Science Program-College of Arts-Imam Abdul Rahman bin Faisal University. *Library Philosophy and Practice*, 1–26.
12. Cheung, S. (2014). Information Security Management for Higher Education Institutions. *Intelligent Data analysis and its Applications, Volume I. Advances in Intelligent Systems and Computing, 297,* 11–19. https://doi.org/10.1007/978-3-319-07776-5_2
13. Al Quhtani, M. (2017). Data Mining Usage in Corporate Information Security: Intrusion Detection Applications. Business Systems Research. *International journal of the Society for Advancing Innovation and Research in Economy, 8(1),* 51–59. https://doi.org/10.1515/bsrj-2017-0005
14. Salem, I., et al. (2022). Introduction to The Data Mining Techniques in Cybersecurity. *Mesopotamian J. Cybersecur. 2022,* 28–37. https://doi.org/10.58496/MJCS/2022/004
15. Kong, J., et al. (2021). Deep-stacking Network Approach by Multisource Data Mining for Hazardous Risk Identification in IoT-based Intelligent Food Management Systems. *Computational Intelligence and Neuroscience, 202*. https://doi.org/10.1155/2021/1194565.
16. Mathew, A. (2023). The Power of Cybersecurity Data Science in Protecting Digital Footprints. *Cognizance J. Multidisciplinary Studies, 3(2),* 1–4. https://doi.org/10.47760/cognizance.2023.v03i02.001
17. Muhammad, S., Dey, B., & Weerakkody, V. (2018). Analysis of Factors That Influence Customers' Willingness to Leave Big Data Digital Footprints on Social Media: A Systematic Review of Literature. *Inf. Syst. Frontiers, 20,* 559–576. https://doi.org/10.1007/s10796-017-9802-y
18. Cheng, F., & Wang, Y. (2018). The do Not Track Mechanism for Digital Footprint Privacy Protection in Marketing Applications. *J. Bus. Econom. Manag. 19(2),* 253–267. https://doi.org/10.3846/jbem.2018.5200

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (23), 2024
ISSN 2663 - 4023

Valeriy Lakhno
Doctor of Technical Sciences, Professor, Department of Computer Systems, Networks and Cybersecurity
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
ORCID 0000-0001-9695-4543
lva964@nubip.edu.ua

Semen Voloshyn
Ph.D, Associate Professor, Associate Professor at the Department of Computer Systems, Networks and Cybersecurity
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
ORCID 0000-0002-4913-7003
voloshyn@nubip.edu.ua

Serhii Mamchenko
Doctor of Technical Sciences, Professor, Department of Computer Systems, Networks and Cybersecurity
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
ORCID 0009-0006-8743-5606
s.mamchenko@nubip.edu.ua

Oleg Kulynich
Ph.D, Associate Professor, Associate Professor at the Department of Computer Systems, Networks and Cybersecurity
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
ORCID 0000-0002-0643-6898
o.kulinich@nubip.edu.ua

Dmytro Kasatkin
Ph.D, Associate Professor, Head of the Department of Computer Systems, Networks and Cybersecurity
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
ORCID 0000-0002-2642-8908
d.kasatkin@nubip.edu.ua

CLUSTER ANALYSIS FOR RESEARCHING DIGITAL FOOTPRINTS OF STUDENTS IN EDUCATIONAL INSTITUTIONS

Abstract. It is shown that Cluster Analysis (CA) can be used in the process of researching the Digital Traces (DT) of students of an educational institution, as well as other educational institutions that introduce a Digital Educational Environment (DEE) into the educational process. Cluster analysis can reveal behavioral patterns of education seekers. Also, the use of CA methods will improve the personalization of training and increase the effectiveness of educational programs. It is shown that in the context of ensuring Information Security (IS) of the DEE of educational institutions, technologies and methods of DT analysis can also be useful, for example, for: monitoring students' network activity; analysis of student authorization and authentication logs; detection of malicious programs and attacks on the DEE; analysis of IS threats to the DEE as a whole; vulnerability prediction. It is shown that the application of CA methods can be useful in studying the degree of information security of the DEE of universities and other educational institutions. It has been established that CA methods can help identify groups of students with similar patterns of activity from the point of view of IS, both the DEE of the educational institution as a whole, and its computer networks and systems. It has been established that with the help of CA DT, it is possible to detect anomalous behavior of students, to detect unusual patterns of activity, facts of unauthorized use of resources or other deviations from the typical behavior of students in the network of the educational institution. The article also provides the results of experimental studies of the level of competences of students of various specialties at the university in IS and protection of information assets of the DEE. In this, CA methods were used in the process of studying students' DT. Six types of users were distinguished on the basis of CA DT of different groups of students registered in the university DEE. As a result of the application of CA methods, students registered in the university's DEE were divided into appropriate clusters according to criteria affecting IS risks.

Keywords digital traces; cluster analysis; digital educational environment of the educational institution; informational security.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Oliveira, P., Cunha, C., & Nakayama, M. (2016). Learning Management Systems (LMS) and E-learning Management: An Integrative Review and Research Agenda. *JISTEM-J. Inf. Syst. Technol. Manag. 13,* 157–180.
2. Aldiab, A., et al. (2019). Utilization of Learning Management Systems (LMSs) in Higher Education System: A Case Review for Saudi Arabia. *Energy Procedia, 160,* 731–737. https://doi.org/10.1016/j.egypro.2019.02.186
3. Azcona, D., Hsiao, I., & Smeaton, A. (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/s11257-019-09234-7
4. Nai, R., et al. (2023). Process Mining on Students' Web Learning Traces: A Case Study with an Ethnographic Analysis. *European Conference on Technology Enhanced Learning, Lecture Note in Computer Science, 14200,* 599–604. https://doi.org/10.1007/978-3-031-42682-7_48
5. Mohssine, B., et al. (2021). Adaptive Help System Based on Learners 'Digital Traces' and Learning Styles. *Int. J. Emerging Technol. Learning (iJET), 16(10),* 288–294. https://doi.org/10.3991/ijet.v16i10.19839
6. Ye, D., & Pennisi, S. (2022). Using Trace Data to Enhance Students' Self-Regulation: A Learning Analytics Perspective. *The Internet and Higher Education, 54, 100855.* https://doi.org/10.1016/j.iheduc.2022.100855
7. Noskova, T., Pavlova, T., & Yakovleva, O. (2018). Study of Students' Educational Activity Strategies in the Social Media Environment. *E-learning and Smart Learning Environment for the Preparation of New Generation Specialists, 10,* 113–125.
8. Kadoić, N., & Oreški, D. (2018). Analysis of Student Behavior and Success Based on Logs in Moodle. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO).* https://doi.org/10.23919/MIPRO.2018.8400123
9. Mogus, A., Djurdjevic, I., & Suvak, N. (2012). The Impact of Student Activity in a Virtual Learning Environment on Their Final Mark. *Active Learning in Higher Education, 13(3),* 177–189. https://doi.org/10.1177/1469787412452985
10. Stiller, K., & Bachmaier, R. (2018). Identifying Learner Types in Distance Training by Using Study Times. *EDEN Conference Proceedings, 1,* 78–86.
11. Ahmed, A., Alharthe, R., & Alfereej, M. (2023). Organizational Committees and Their Role in Enhancing Intellectual Security: A Case Study on Female Students of the Bachelor of Information Science Program-College of Arts-Imam Abdul Rahman bin Faisal University. *Library Philosophy and Practice,* 1–26.
12. Cheung, S. (2014). Information Security Management for Higher Education Institutions. *Intelligent Data analysis and its Applications, Volume I. Advances in Intelligent Systems and Computing, 297,* 11–19. https://doi.org/10.1007/978-3-319-07776-5_2
13. Al Quhtani, M. (2017). Data Mining Usage in Corporate Information Security: Intrusion Detection Applications. Business Systems Research. *International journal of the Society for Advancing Innovation and Research in Economy, 8(1),* 51–59. https://doi.org/10.1515/bsrj-2017-0005
14. Salem, I., et al. (2022). Introduction to The Data Mining Techniques in Cybersecurity. *Mesopotamian J. Cybersecur. 2022,* 28–37. https://doi.org/10.58496/MJCS/2022/004
15. Kong, J., et al. (2021). Deep-stacking Network Approach by Multisource Data Mining for Hazardous Risk Identification in IoT-based Intelligent Food Management Systems. *Computational Intelligence and Neuroscience, 202.* https://doi.org/10.1155/2021/1194565.
16. Mathew, A. (2023). The Power of Cybersecurity Data Science in Protecting Digital Footprints. *Cognizance J. Multidisciplinary Studies, 3(2),* 1–4. https://doi.org/10.47760/cognizance.2023.v03i02.001
17. Muhammad, S., Dey, B., & Weerakkody, V. (2018). Analysis of Factors That Influence Customers' Willingness to Leave Big Data Digital Footprints on Social Media: A Systematic Review of Literature. *Inf. Syst. Frontiers, 20,* 559–576. https://doi.org/10.1007/s10796-017-9802-y
18. Cheng, F., & Wang, Y. (2018). The do Not Track Mechanism for Digital Footprint Privacy Protection in Marketing Applications. *J. Bus. Econom. Manag. 19(2),* 253–267. https://doi.org/10.3846/jbem.2018.5200