КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (27), 2025
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2025.27.726
УДК 004.056.5:004.7

Щавінський Юрій Віталійович
к.т.н., доцент, доцент кафедри управління кібербезпекою та захистом інформації
Державний університет інформаційно-комунікаційних технологій, Київ, Україна
ORCID ID: 0000-0002-2319-8983
yushchavinsky@ukr.net

Будзинський Олександр Володимирович
аспірант кафедри управління кібербезпекою та захистом інформації
Державний університет інформаційно-комунікаційних технологій, Київ, Україна
ORCID ID: 0009-0002-2402-0711
oleksandr.email@gmail.com

АНАЛІЗ АКТУАЛЬНИХ ПРОБЛЕМ БЕЗПЕКИ КОРПОРАТИВНИХ БАЗ ДАНИХ В УМОВАХ СУЧАСНОЇ ІНФРАСТРУКТУРИ ТА ШЛЯХИ ЇХ ВИРІШЕННЯ

Анотація. Дослідження спрямоване на аналіз актуальних проблем безпеки корпоративних баз даних в умовах сучасної інфраструктури, розробку моделі виявлення аномальної активності доступу до баз даних та інтеграцію її в SIEM-систему AlienVault для автоматичного реагування на загрози. Встановлено, що одною із головних проблем захисту баз даних є потреба в негайному виявленні аномалій доступу та реагуванні на загрозиконфіденційності, доступності та цілісності баз даних. Аналіз наукової літератури дозволив зробити висновок, що сучасна інфраструктура значно змінює підхід до безпеки корпоративних баз даних, створюючи як нові можливості, так і загрози. Виклики, повязані із впливом сучасної інфраструктури на безпеку баз даних потребують нових шляхів вирішення проблем та проактивного комплексного підходу, який полягає у застосуванні штучного інтелекту при організації захисту корпоративних баз даних У дослідженні застосований алгоритм IsolationForest для створення моделі виявлення аномалій доступу до баз даних. з використанням бібліотек вільного доступу мови програмування Python. Проведено навчання моделі на основі історичних даних з подальшим тестуванням та оцінкою ефективності за метриками Accuracy, Precision, Recall, ROC AUC. В процесі навчання моделі досягнуто високої точності виявлення аномалій (Accuracy = 98,8%, ROC AUC ≈ 0,99, Precision = 0,86, Recall = 0,99). Інтеграцію моделі реалізовано через механізм виклику зовнішніх скриптів у AlienVault. Розроблена модель дозволяє ідентифікувати потенційні загрози в режимі реального часу, оцінювати рівень ризику та автоматично блокувати небезпечні запити або надсилати сповіщення в SOC. Запропоновано підхід інтеграції алгоритму машинного навчання в SIEM-системи, що забезпечує проактивне виявлення та реагування на загрози безпеки баз даних. Впровадження даної моделі дозволяє підвищити рівень безпеки корпоративних баз даних, зменшити ризики витоку інформації та забезпечити оперативне реагування на кіберінциденти.

Ключові слова: кібербезпека; безпека баз даних; машинне навчання; SIEM; AlienVault; IsolationForest; SOC.

ВСТУП

Сьогодні інформаційні технології є основою функціонування урядових структур, організацій та бізнесу. Сучасна мережева інфраструктура все більше орієнтована на використання хмарних платформ, таких як AWS, Microsoft Azure та Google Cloud. Це дозволяє компаніям масштабувати свої ресурси, забезпечуючи гнучкість та ефективність управління даними. Проте хмарні рішення несуть нові ризики, пов’язані з витоками даних, несанкціонованим доступом та неправильними налаштуваннями конфігурацій.

Зростання використання IoT-пристроїв та мобільних девайсів у корпоративному середовищі відкриває нові вектори атак [1]. Відсутність надійної сегментації мережі та незахищені пристрої можуть бути використані як точки входу для атак на корпоративні бази даних. Використання Mobile Device Management (MDM), сегментація мережі та ізоляція IoT-пристроїв є необхідними заходами для зменшення цих ризиків. Сучасна інфраструктура базується на автоматизованих процесах CI/CD (Continuous Integration/Continuous Deployment), що забезпечує швидке оновлення додатків. Однак відсутність контролю безпеки в цих процесах може призвести до розгортання вразливих версій програмного забезпечення.

Корпоративні бази даних, які служать сховищами для найбільш важливих і цінних даних компанії, містять конфіденційну інформацію, яка є критично важливою для їхньої діяльності. Зростання популярності систем БД супроводжується сплеском інцидентів витоку даних. З розвитком мережевих технологій та змін у способах зберігання й обробки даних, питання безпеки корпоративних баз даних стає все більш актуальним. Розподілена інфраструктура, що включає кілька хмарних сервісів і дата-центрів, ускладнює контроль за доступом до корпоративних даних. Використання API (Application Programming Interface) для інтеграції різних сервісів створює потенційні вразливості, які можуть бути використані зловмисниками.

Людський фактор залишається одним із найбільших викликів у забезпеченні безпеки корпоративних БД. Фішингові атаки, недостатня автентифікація користувачів та інсайдерські загрози можуть призвести до витоку або компрометації даних.

Постановка проблеми. Таким чином, сучасна мережева інфраструктура змінює традиційні підходи до безпеки корпоративних баз даних, роблячи її більш гнучкою, але водночас складнішою для захисту, що вимагає підвищеного рівня пильності по відношенню до інших систем в компанії. Для підтримки цілісності БД необхідний репертуар комплексних заходів захисту БД. Виклики, пов’язані з хмарними технологіями, IoT, DevOps, API та соціальною інженерією, вимагають проведення досліджень з метою розроблення та впровадження передових заходів безпеки.

Аналіз останніх досліджень і публікацій. Розв’язанням актуальних проблем захисту корпоративних баз даних в умовах сучасної інфраструктури займалась велика кількість вітчизняних та зарубіжних науковців. В дослідженні [2] наведено інформацію щодо основних загроз безпеці інформації і вразливих місць ІТ-систем та обґрунтована необхідність проведення пентесту в ІТ-системах (мережах) органів влади та різних державних реєстрів задля об’єктивної оцінки рівня безпеки цих структур в умовах сучасної інформаційної та кібервійни, яка ведеться проти нашої країни. Особливо важливим постало питання захисту баз даних урядових структур у зв’язку з наймасштабнішою кібератакою хакерів на державні реєстри України, що привело до порушення роботи критично важливої інфраструктури держави.

В роботах [3] – [7] науковцями оцінені стратегії, що використовуються для посилення безпеки даних, особлива увага приділена методам шифрування, контролю доступу та механізмам автентифікації, досліджені нові тенденції в безпеці даних, такі як інтеграція штучного інтелекту (ШІ) та машинного навчання для розширеного виявлення загроз, потенціал блокчейну для підвищення цілісності даних та прийняття архітектур з нульовою довірою.

Автори в роботах [8] – [11] розглянувши різні проблеми безпеки БД, встановили, що безпека великомасштабних баз даних значною мірою залежить від різноманітних захисних механізмів, у дослідженні [12] науковці запропонували інноваційну методику виявлення різних загроз для систем БД шляхом оцінки ризику для нових вхідних активностей. Їх дослідження виявило різні шкідливі атаки, які можуть завдати шкоди системі БД. При цьому, акцент у їхніх дослідженнях обмежується лише оцінкою безпеки за участю БД.

Підхід у дослідженні [13], який оснований на оцінці ризиків, використовує функції рольової поведінки користувача для виявлення нав'язливої атаки. У дослідженні встановлено, що у системі виявлення вторгнень у БД для середовища доступу на основі ролей (RBAC — Role-Based Access Control) недостатньо зосередитися на функціях, заснованих на ролях, оскільки кожен користувач у межах однієї ролі має певний ступінь унікальності.

Аналіз публікацій показує, що безліч досліджень заглибилися в різні аспекти безпеки БД, наукова спільнота з безпеки БД розробила низку стратегій і процедур для захисту конфіденційності, цілісності та доступності інформації, що зберігається в корпоративних БД. Разом з тим, потрібно зауважити, що впливу сучасної мережевої інфраструктури на стан безпеки БД уваги приділено недостатньо.

Метою статті є аналіз сучасного стану безпеки корпоративних баз даних в умовах сучасної інфраструктури та визначення шляхів вирішення проблем захисту баз даних.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Аналіз впливу сучасної інфраструктури на безпеку корпоративних баз даних

Сучасна мережева інфраструктура зазнала значних змін, які вплинули на організацію, зберігання та захист корпоративних БД. Перехід до хмарних технологій, розподілених обчислень, віртуалізації та використання ШІ відкриває нові можливості, але водночас створює нові загрози та виклики [14].

Розташування БД компанії в середовищі, яке контролюється хмарним провайдером підвищує ризик несанкціонованого доступу та ускладнює контроль над ними у зв'язку з неможливістю чіткого визначення місця знаходження даних. А використання загальнодоступних інтерфейсів може призвести до атак, таких як перехоплення токенів доступу чи DDoS-атаки на хмарні сервіси. Розподілення між декількома дата-центрами чи хмарними платформами створює додаткові виклики, пов'язані із складнощами визначення політик доступу для різних користувачів і пристроїв. Розподілена інфраструктура вимагає складних механізмів синхронізації та реплікації, що підвищує ризик помилок.

Зростаюча кількість підключених IoT-пристроїв, які сьогодні мають недостатній рівень захисту, приводить до зростання поверхні атак та створює навантаження на мережеву інфраструктуру, що може впливати на продуктивність та безпеку БД.

Штучний інтелект використовується не тільки для покращення безпеки, використання його зловмисниками для аналізу вразливостей та проведення атак з високою точністю, алгоритмів глибокого навчання для створення персоналізованих атак потребує значних затрат та вдосконалених криптографічних механізмів. Порівняльний аналіз впливу сучасної мережевої інфраструктури наведений в табл. 1.

Таблиця 1

Порівняльний аналіз впливу сучасної інфраструктури на безпеку корпоративних баз даних

Технологія/Фактор	Позитивний вплив	Негативний вплив
Хмарні обчислення	Зниження витрат, гнучкість, масштабованість	Ризики доступу третіх сторін, залежність від провайдера
Розподілені системи	Підвищена надійність, географічна розподіленість	Проблеми синхронізації, складність управління
IoT-пристрої	Автоматизація, нові можливості аналізу	Недостатній рівень захисту, зростання поверхні атаки
Штучний інтелект	Виявлення аномалій, автоматизація захисту	Автоматизовані атаки, складність налаштування
Криптографія	Високий рівень безпеки, захист переданих і збережених даних	Витрати на обчислення, складність впровадження

Аналіз табл. 1 дозволяє зробити висновок, що сучасна інфраструктура значно змінює підхід до безпеки корпоративних БД, створюючи як нові можливості, так і загрози. Для ефективного захисту необхідно впроваджувати багаторівневі механізми безпеки, використовуючи поєднання хмарних технологій, ШІ, криптографічних рішень та систем управління доступом. Важливим фактором залишається проактивний підхід, що включає безперервний моніторинг, аналіз загроз та швидке реагування на інциденти. Лише такий комплексний підхід дозволить забезпечити належний рівень захисту корпоративних даних у сучасних умовах цифрової трансформації.

Шляхи вирішення проблем безпеки корпоративних баз даних в умовах сучасної мережевої інфраструктури

Проблеми безпеки в БД викликали необхідність розробки рішень цих проблем. Найкраща інфраструктура інформаційної безпеки не може гарантувати, що вторгнення або інші зловмисні дії не відбудуться. Проактивний підхід до безпеки корпоративних БД означає випереджувальне реагування на потенційні загрози замість того, щоб діяти лише після того, як інцидент уже стався. Традиційні методи безпеки, не зважаючи на певну ефективність, часто реагують на інциденти після їх виникнення. Застосування ШІ представляє проактивні можливості, такі як прогнозна аналітика, виявлення аномалій, системи автоматичного реагування, які значно покращують безпеку БД. Прогностична аналітика на основі ШІ дозволяє організаціям прогнозувати потенційні загрози на основі історичних моделей даних і тенденцій. Це дає змогу здійснювати попереджувальні дії, які зменшують уразливості, перш ніж їх можна буде застосувати зловмисникам. Основні принципи проактивного підходу визначені в табл. 2.

Таблиця 2

Основні принципи проактивного підходу

Принципи	Механізми
Моніторинг у реальному часі та виявлення аномалій	Використання SIEM-систем (QRadar, AlienVault) для аналізу логів та виявлення підозрілої активності. Використання поведінкового аналізу для виявлення аномальних дій користувачів та процесів. Інтеграція штучного інтелекту для розпізнавання складних атак, включаючи 0-day експлойти.
Проактивне тестування безпеки	Пентестинг (penetration testing) для ідентифікації вразливостей. Red Team/Blue Team підхід, коли одна група імітує атаки, а інша захищається. Використання автоматизованих сканерів вразливостей (Nessus, OpenVAS).

Моделювання загроз та оцінка ризиків	Аналіз можливих векторів атак за методологією MITRE ATT&CK. Використання Threat Intelligence (наприклад, AlienVault OTX) для отримання даних про актуальні загрози. Побудова моделей загроз для критичних систем та баз даних.
Жорстке управління доступом та контроль привілеїв	Використання RBAC (Role-Based Access Control) або ABAC (Attribute-Based Access Control). Реалізація принципу найменших привілеїв (Least Privilege). Багатофакторна автентифікація (MFA) для захисту адміністративних облікових записів.
Безперервне оновлення та патчинг	Регулярне оновлення ПЗ, ОС та СУБД для усунення відомих вразливостей. Використання механізмів автоматичного патчингу та віртуальних патчів для критичних сервісів.
Аварійне відновлення	Розробка планів реагування на інциденти (Incident Response Plan). Використання резервного копіювання з можливістю швидкого відновлення (Backup & Disaster Recovery). Тестування планів BCP (Business Continuity Planning).

Аналіз табл. 2 дозволяє зробити висновок про те, що реалізація цих принципів вимагає комплексного підходу, який повинен включати технології на основі ШІ, процеси та підготовку персоналу.

Проактивний підхід до захисту корпоративних БД починається з виявлення та оцінки потенційних загроз, що дозволяє визначити можливі вразливості та ризики, які можуть вплинути на БД. Сьогодні дослідниками запропоновано багато методів вирішення проблем, пов’язаних із застосуванням ШІ, який відіграє важливу роль у вдосконаленні можливостей виявлення загроз, покращенні стратегій реагування та підвищенні рівня виявлення аномалій у системах БД. Крім того, інтеграція ШІ посилює безперервний моніторинг, дозволяючи командам безпеки підтримувати цілодобову пильність над системами БД. Ці системи на основі ШІ розвиваються з часом, навчаючись на нових даних і підвищуючи свою точність у виявленні нових загроз. Вони також допомагають організаціям дотримуватися правил захисту даних, надаючи журнали аудиту та гарантуючи, що бази даних відповідають стандартам безпеки. Сучасні системи захисту інформації на основі ШІ вимагають переходу від реактивних методів до проактивного підходу, що забезпечує випередження загроз задовго до їх активної фази. Однак у впровадженні рішень ШІ залишаються проблеми, зокрема керування хибними спрацьовуваннями, підтримка якості даних і забезпечення ефективної інтеграції з існуючою інфраструктурою безпеки [18].

Виявлення аномалій на основі машинного навчання — це позначення незвичайних дій чи поведінки в режимі реального часу, що дозволяє виявляти складні загрози, які можуть обійти звичайні заходи безпеки. Виявлення аномалій використовує алгоритми машинного навчання для встановлення базової лінії нормальної поведінки в системі. Ця базова лінія побудована на основі історичних даних, включаючи типові дії користувачів, моделі доступу до даних і мережевий трафік. Після того, як система дізналася, що є «нормальною» поведінкою, вона постійно відстежує дії в реальному часі та позначає будь-які відхилення від цих норм. Наприклад, якщо користувач зазвичай отримує доступ до певного набору даних у робочий час і раптом починає завантажувати великі обсяги конфіденційних даних у непарні години, це буде позначено як аномалія. Подібним чином, якщо в базі даних раптово спостерігається сплеск трафіку з незнайомої IP-адреси, системи виявлення аномалій попередять команду безпеки про потенційне вторгнення. Ці та інші сповіщенняпід час роботи створених моделей на основі ШІ в реальному часі мають вирішальне значення для запобігання витоку даних до їх ескалації.

Одною з таких моделей є модель IsolationForest, як перший етап у ланцюгу проактивного захисту та ефективний інструмент для виявлення аномалій в журналах доступу до баз даних [15]. IsolationForest є алгоритмом, що працює на основі принципу ізоляції даних. Його основна ідея полягає в тому, що аномальні дані легше ізолювати, ніж нормальні, завдяки їхній відмінності від загальної маси. Алгоритм будує набір дерев (ізоляційний ліс), де кожне дерево випадково обирає ознаки та порогові значення для поділу даних. Об'єкти, що потребують меншої кількості розділень для ізоляції, класифікуються як аномалії.

Принципова схема побудови та роботи моделі зображена на рис. 1.

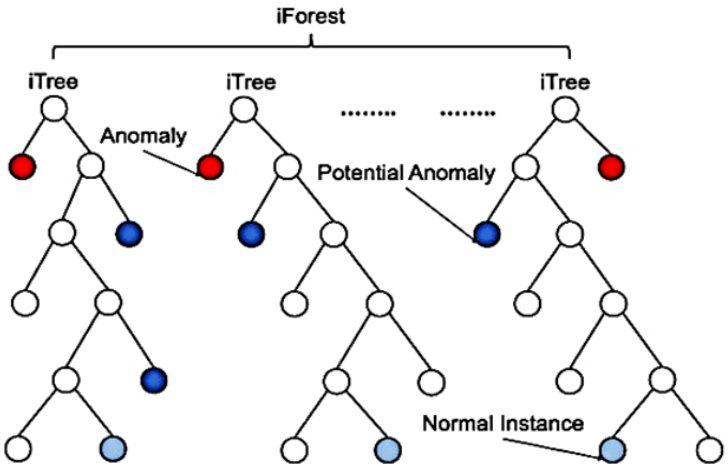

Рис. 1. Принцип роботи моделі IsolationForest

Застосування моделі дозволяє в роботі з великими наборами даних і різнорідними ознаками, що є характерним для журналів доступу до баз даних, швидко та точно виділяти підозрілі записи, які рідко зустрічаються у звичайних сценаріях.

Застосування IsolationForest для аналізу журналів доступу до БД дозволяє випереджати потенційні атаки ще на стадії їх початку. Модель може слугувати першим шаром фільтрації, виявляючи підозрілі поведінкові патерни до того, як загроза стане актуальною.

Враховуючи ефективність моделі та наявність бібліотек для розробки і готових даних для навчання, під час дослідження розроблена модель з використанням об'єктно орієнтованої мови програмування високого рівня загального призначення з відкритим кодом Python для виявлення аномалій доступу до корпоративних БД на основі логфайлів та впроваджена в SIEM AlienVault за алгоритмом, який зображений на рис. 2.

При налаштуванні правил створення логфайла в AlienVault зміст записів у цей файл включав:

- дату і час доступу до БД;
- IP-адресу та порт доступу до БД;
- IP-адресу і порт звідки здійснювався запит;
- тип запиту і його довжина;
- тривалість доступу до БД;
- кількість змінених записів в БД;
- кількість спроб введення паролю доступу до БД;
- зміна привілеїв доступу.

ISSN 2663 - 4023

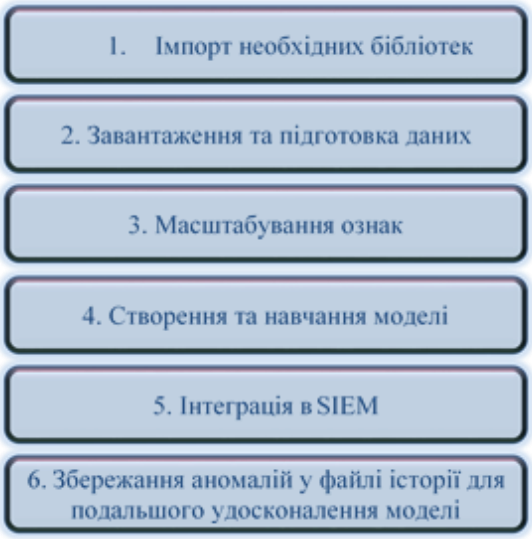

Рис. 2. Алгоритм створення моделі

З метою навчання моделі використаний запропонований розроблений науковцями у своєму дослідженні [16] навчальний набір, який є у вільному доступі та найбільше відповідає потребам виявлення аномалій у відповідності із структурою логфайлів AlienVault. Загалом, цей набір даних є цінним ресурсом для дослідників, аналітиків і розробників, які працюють над рішеннями з кібербезпеки. Він забезпечує багату, різноманітну та реалістичну основу для розробки та тестування моделей машинного навчання, спрямованих на виявлення та пом'якшення кібератак. Комплексні атрибути та реалістичні недоліки роблять його ідеальним інструментом для практичних реальних застосувань у дослідженнях і розробках кібербезпеки.

При створенні моделі використані доступні бібліотеки, наведені у фрагменті коду:

```
import pandas as pd
import numpy as np
from sklearn.ensemble import IsolationForest
from sklearn.metrics import accuracy_score,
roc_auc_score, precision_recall_curve, confusion_matrix, roc_curve
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import roc_curve, precision_recall_curve,
confusion_matrix, accuracy_score, roc_auc_score, ConfusionMatrixDisplay
import os
import csv
from sklearn.model_selection import train_test_split
```

Використання навчального набору дозволило навчити модель, вхідними параметрами якої є *n_estimators, contamination, random_state*, які протягом навчання можена змінювати:

```
# Створення та навчання моделі Isolation Forest
iforest = IsolationForest(contamination=0.1,random_state=42, n_estimators=300)
iforest.fit(X_train)
```

Параметр *contamination* у моделі IsolationForest визначає частку аномальних точок у наборі даних. Від цього параметра залежить, як модель визначатиме поріг для класифікації точок як аномалій. Вищий рівень contamination означає, що модель очікує більше аномалій у даних, і відповідно поріг для класифікації точки як аномалії буде нижчим. Якщо contamination встановлено занадто низьким або занадто високим, це може призвести до помилкових позитивних або негативних результатів. Параметр contamination впливає на процес навчання моделі, оскільки він визначає, які точки будуть використовуватися для побудови моделі. Тому важливо налаштувати цей параметр відповідно до реальної частки аномалій у даних.

Параметр *n_estimators* у моделі IsolationForest визначає кількість дерев у лісі. Збільшення кількості дерев може підвищити точність моделі, оскільки більше дерев дозволяє моделі краще виявляти аномалії. Разом з тим, більша кількість дерев збільшує час, необхідний для навчання моделі. Це може бути важливо, якщо є великий набір даних або обмежені обчислювальні ресурси. Але більша кількість дерев може зробити результати моделі більш стабільними та менш залежними від випадкових варіацій у даних.

Параметр *random_state* у моделі IsolationForest впливає на відтворюваність результатів. Встановлення значення random_state дозволяє отримувати однакові результати при кожному запуску моделі з тими ж даними та параметрами. Параметр random_state контролює початкові значення для випадкових процесів, таких як розбиття даних на тренувальні та тестові набори або вибір випадкових підмножин даних для побудови дерев у лісі.

Ознаками для навчання моделі з навчального набору вибрані порт доступу до бази даних, IP-адреса та порт звідки здійснено запит, тип запиту і його довжина, час запиту, кількість змінених записів у базі даних, тривалість сесії та зміна привілеїв доступу:

```
# Вибір ознак для навчання
features = ['src_port', 'dst_IP', 'dst_port',
'query_length', 'response_time',
'affected_rows','failed_login_attempts',
'session_duration','privilege_escalation']
```

Результати моделювання з різними варіаціями параметрів моделі під час навчання та оцінювання якості на контрольному (тестовому) наборі наведені в табл. 3 та відображені на рис. 3.

Таблиця 3

Результати моделювання

кількість дерев (n_estimators)	contamination							
	0,01		0,05		0,1		0,15	
	Accuracy	ROC AUC	Accuracy	ROC AUC	Accuracy	ROC AUC	Accuracy	ROC AUC
100	0,9382	0,542	0,9787	0,8465	0,9384	0,5428	0,942	0,5719
200	0,939	0,5424	0,9795	0,8521	***0,9782***	***0,984***	0,9155	0,9546
300	0,9385	0,5461	0,9815	0,8634	***0,969***	***0,9833***	0,913	0,9533
400	0,9385	0,5461	0,9887	0,8579	***0,9672***	***0,9824***	0,917	0,9554
500	0,9385	0,5461	0,981	0,8597	0,966	0,9817	0,9142	0,954

ISSN 2663 - 4023

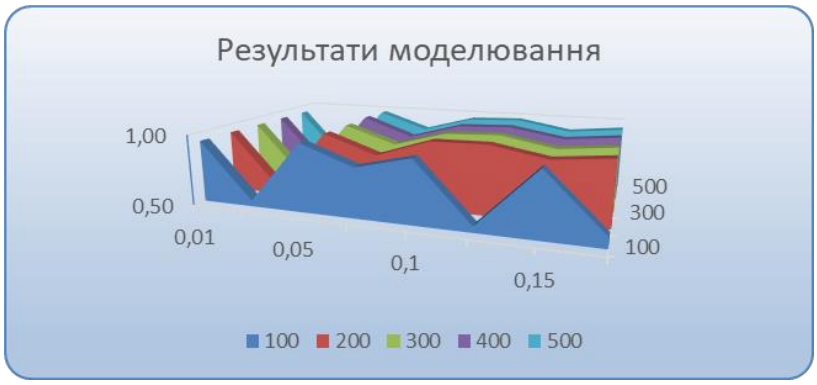

Рис. 3. Результати навчання моделі IsolationForest

Аналіз табл. 3 та рис. 3 дозволяє зробити остаточний висновок, що кращими параметрами моделі є contamination = 0,1 і n_estimators = {200,300,400}. Подальше збільшення цих параметрів приводить до зменшення точності визначення аномалій, а збільшення кількості дерев, крім того, до збільшення часу реакції моделі.

Створена модель майже ідеально визначила аномалії в тестовому наборі даних. При подальшому моделюванні, використовуючи метод інтерполяції найкращих показників (Accuracy: 0.98875, ROC AUC: 0.9922552765350663) модель навчилась при параметрах *contamination* = 0,08 і *n_estimators* = 400. В цьому варіанті створена модель майже ідеально визначила аномалії в тестовому наборі даних. Результати ідентифікації правильного та хибного визначення аномалій на тестовому наборі зосереджені в матриці невідповідностей та показані на рис. 4.

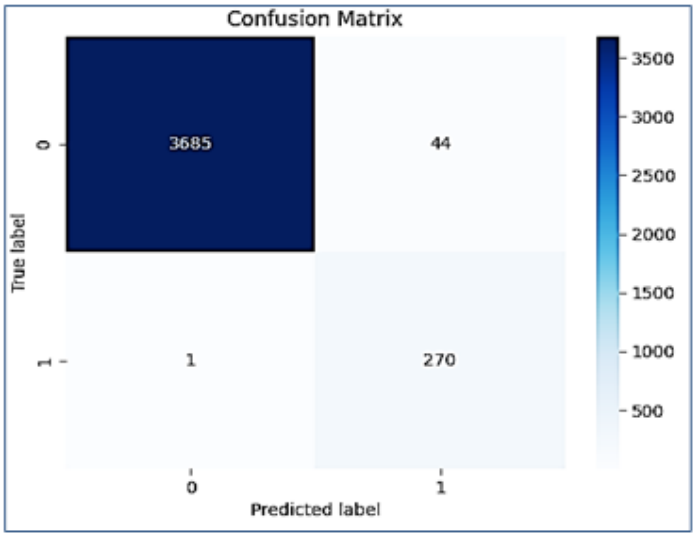

Рис. 4. Матриця невідповідностей

True Positive (TP) = 270 — аномальні запити, які модель правильно визначила як аномалії.

True Negative (TN) = 3685 — нормальні запити, які модель правильно ідентифікувала як нормальні.

False Positive (FP) = 44 — нормальні запити, які модель помилково класифікувала як аномальні.

False Negative (FN) = 1 — аномальний запит, який модель не виявила.

ISSN 2663 - 4023

На рис. 5 показані співвідношення основних метрик якості навчання моделі з використанням матриці невідповідності (рис. 4).

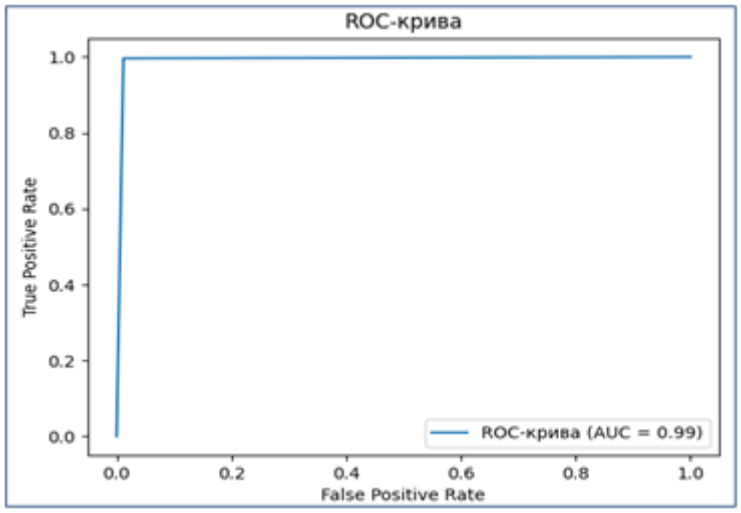

Рис. 5. Результат визначення ROS-кривої та AUS

True Positive Rate (TPR) або чутливість (Recall) відображає частку правильно передбачених позитивних випадків серед усіх фактичних позитивних зразків [17]

$$Recall = TPR = \frac{TP}{TP+FN}, \quad (1)$$

де TP — кількість аномальних запитів, які модель правильно визначила як аномалії, FN — кількість аномальних запитів, яких модель не виявила.

False Positive Rate (FPR) визначає, скільки негативних випадків помилково класифіковані як позитивні [17]

$$FPR = \frac{FP}{FP+TN}, \quad (2)$$

де FP — кількість нормальних запитів, які модель помилково визначила як аномалії, TN — кількість нормальних запитів, яких модель правильно виявила.

Площа під кривою AUC (Area Under Curve) визначена за формулою [17]

$$AUS = \int_0^1 TPRdFPR \quad (3)$$

Оціночна метрика *Precision* показує відсоток реальних аномальних запитів серед усіх, які модель позначила як аномалії, і визначається за формулою [17]

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

Ймовірність, що запити, визначені як нормальні, дійсно є нормальними, розраховується за формулою [17]

$$NPR = \frac{TN}{TN + FN} \quad (5)$$

Ще одною оціночною метрикою моделі є *F1-score*, яка відображає баланс між *Precision* (точністю) та *Recall* (повнотою) у задачах класифікації. Якщо F1 близька до 1, то модель добре ідентифікує позитивні зразки, не допускаючи багато помилок. Якщо F1 низька, то модель або пропускає багато позитивних випадків, або робить надто багато хибнопозитивних прогнозів. Дана метрика розраховується за формулою [17]

$$F1 = 2\,\frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

Таким чином, використовуючи формул (1) – (6) для оцінювання створеної та навченої моделі, за результатами моделювання отримали наступні показники якості:

- Recall = 0.996 — модель має високу повноту, що означає, що вона майже не пропускає аномалії;
- Precision = 0.86, що означає, що приблизно 86% позначених аномалій є реальними загрозами;
- F1-score = 0.92 вказує на баланс між Precision і Recall;
- FPR = 0.012 — дуже низький показник хибних тривог;
- NPR ≈ 0.9997…, що свідчить про точність визначення нормальних запитів;
- AUC = 0.99 означає, що модель майже ідеальна у розподілі позитивних та негативних випадків.

Наступним етапом запропонованого алгоритму застосування моделі виявлення аномалій доступу є впровадження навченої моделі в SIEM AlienVault. Загальний підхід до інтеграції полягає у виконанні визначеної послідовності кроків (рис. 6).

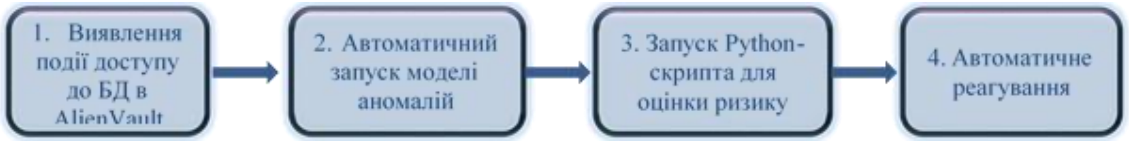

Рис. 6. Алгоритм впровадження моделі

1. Для виявлення події доступу до БД в AlienVault необхідно налаштувати правила для виявлення цих подій через налаштування плагіна моніторингу доступу до БД, відслідковуючи запити

```
[DEFAULT]
plugin_id=9002
type=detector
enable=yes

[config]
source=log
location=/var/log/db_access.log
process=my_db_plugin

[translation]
log=event

[event]
event_type=db_access
regex=(?P<date>\d{4}-\d{2}-\d{2} \d{2}:\d{2}:\d{2}) - (?P<user>\w+) -
(?P<src_ip>\d+\.\d+\.\d+\.\d+) - (?P<query>.+)
```

2. Коли AlienVault виявляє запит, необхідно автоматично запускати Python-скрипт для аналізу через створення скрипта запуску моделі.

```
sudo nano /usr/local/bin/anomaly_detector.sh
```

та заповнивши його

```
/usr/bin/python3 /usr/local/bin/anomaly_model.py "$1" "$2" "$3"
```

3. Оцінювання виявленого аномального доступу (ризику) з логуванням результату для подальшого удосконалення моделі в наступному фрагменті коду:

```
# Оцінка ризику (аномалії)
risk_score = model.decision_function([features])[0]
anomality = abs(risk_score)

# Логування результату
log_file = "/var/log/anomaly_detector.log"
logging.basicConfig(filename=log_file, level=logging.INFO, format="%(asctime)s
- %(message)s")
log_message = f"{timestamp} - {user} - {src_ip} - {query} - Risk:
{anomality:.2f}"
logging.info(log_message)
```

4. Реагування на визначені аномалі визначаються рівнями ризику, які можна редагувати. Якщо рівень більше 0,5 — необхідно заблокувати доступ і надіслати сповіщення SOC, якщо менше — тільки надіслати сповіщення.

```
# Реагування
if anomality > 0.5:
    # Блокування IP
    os.system(f"sudo iptables -A INPUT -s {src_ip} -j DROP")
    os.system(f"echo '{src_ip} blocked due to high anomaly risk
    {anomality:.2f}' | mail -s 'DB Threat Alert' soc@company.com")
    print(f"Blocked {src_ip} due to risk {anomality:.2f}")
else:
    print(f"Logged event {query} with risk {anomality:.2f}")
    print(f"Сповіщення адміністратора: Аномалія виявлена у користувача
{row['user']} з IP-адресою {row['ip_address']} у країні {row['location']}")
```

Таким чином, розроблена модель забезпечить аналіз доступу до бази даних у реальному часі, виявляючи аномальну активність, та автоматично реагуючи на потенційні загрози. Вона у поєднанні з SIEM аналізує кожен запит до БД на наявність аномалій, визначає потенційні SQL-ін’єкції, спроби підбору паролів та нестандартну активність, при ризику >0.5 блокує доступ, при меншій загрозі надсилає сповіщення в SOC, формує базу виявлених загроз для подальшого навчання та вдосконалення моделі.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Забезпечення безпеки корпоративних баз даних в умовах сучасної мережевої інфраструктури є складним, але необхідним завданням. Виявлено основні загрози безпеці корпоративних БД, які потребують проактивного підходу до їх захисту на основі ШІ.

Одним із шляхі проактичного підходу до вирішення проблем захисту корпоративних БД є розроблення комплексної системи виявлення аномалій доступу до БД та оперативне реагування на загрози. Основою комплексної системи та першим кроком її реалізації є розроблення моделей машинного навчання з використанням мови програмування. Розроблена модель на основі відкритих бібліотек Python за алгоритмом IsolationForest в процесі навчання показала майже ідеальні результати якості за параметрами Accuracy = 98,8%, ROC AUC ≈ 0,99, Precision = 0,86, Recall = 0,99, що дає змогу стверджувати про її придатність до практичної інтеграції в системи безпеки корпоративних БД. Вона дозволяє виявляти потенційні загрози у реальному часі та автоматично блокувати небезпечні дії. Реалізовано механізм оцінки рівня загрози, який

дозволяє здійснювати динамічне реагування — блокування доступу або надсилання сповіщень у SOC. Впровадження автоматизованої моделі аналізу аномалій дозволяє підвищити рівень безпеки корпоративних баз даних та значно зменшити ризик витоку інформації.

Подальші дослідження будуть спрямовані на вдосконалення методів аналізу та автоматизацію заходів реагування, що сприятиме підвищенню рівня безпеки корпоративних баз даних в умовах динамічних кіберзагроз:

- оптимізація алгоритму для покращення швидкодії моделі для ще швидшого реагування на загрози;
- розширення джерел даних, інтеграція додаткових лог-файлів, поведінкових факторів та SIEM-аналітики для підвищення точності прогнозування загроз;
- впровадження адаптивного навчання з метою автоматичного оновлення моделі на основі нових виявлених загроз, що забезпечить її актуальність;
- розробка модуля глибокого аналізу атак для детальної класифікації аномальних запитів для більш ефективної обробки інцидентів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Kostiuk, Y., Bebeshko, B., Kriuchkova, L., Lytvynov, V., Oksanych, I., Skladannyi, P., & Khorolska, K. (2024). Information protection and data exchange security in wireless mobile networks with authentication and key exchange protocols. *Cybersecurity: Education, Science, Technique, 1(25),* 229–252. https://doi.org/10.28925/2663-4023.2024.25.229252
2. Kyrychok, R. V., Skladannyi, P. M., Buryachok, V. L., Hulak, G. M., & Kozachok, V. A. (2016). Problems of ensuring control over the security of corporate networks and ways to solve them. *Scientific Notes of the Ukrainian Research Institute of Communications, 3(43),* 48–61.
3. Adenubi, A. O., & P. Oduroye, A. (2024). Data security in big data: challenges, strategies, and future trends. *International journal of research in education humanities and commerce, 05(02),* 01–15. https://doi.org/10.37602/ijrehc.2024.5201
4. Li, X., Wang, Z., Leung, V. C. M., Ji, H., Liu, Y., & Zhang, H. (2021). Blockchain-empowered data-driven networks. *ACM Computing Surveys, 54(3),* 1–38. https://doi.org/10.1145/3446373
5. Bao, R., Chen, Z., & Obaidat, M. S. (2018). Challenges and techniques in Big data security and privacy: A review. *Security and Privacy, 1(4),* Article e13. https://doi.org/10.1002/spy2.13
6. Yeoh, W., Liu, M., Shore, M., & Jiang, F. (2023). Zero trust cybersecurity: Critical success factors and a maturity assessment framework. *Computers & Security, 103412.* https://doi.org/10.1016/j.cose.2023.103412
7. Prince, N. U., Faheem, M. A., Khan, O., Hossain, K., Alkhayyat, A., Hamdache, A., & Elmouki, I. (2024). AI-Powered data-driven cybersecurity techniques: boosting threat identification and reaction. *Nanotechnology Perceptions 20(S10),* 332–353. https://doi.org/10.62441/nano-ntp.v20is10.25
8. Paul, P., & Aithal, P. S. (2019). Database Security: An overview and analysis of current trend. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3497728
9. Yasmeen, M. (2018). NOSQL database engines for big data management. *International Journal of Trend in Scientific Research and Development, 2(6),* 617–622. https://doi.org/10.31142/ijtsrd18608
10. Mousa, A., Karabatak, M., & Mustafa, T. (2020). Database security threats and challenges. *8th international symposium on digital forensics and security (ISDFS).* https://doi.org/10.1109/isdfs49300.2020.9116436
11. Zaw, T. M., Thant, M., & Bezzateev, S. V. (2019). Database security with AES encryption, elliptic curve encryption and signature. *2019 wave electronics and its application in information and telecommunication systems (WECONF).* https://doi.org/10.1109/weconf.2019.8840125
12. Mataz, A., & Haripriya, V. (2024). Enhancing data protection through advanced encryption or improving data security with advanced encryption. *International Journal of Innovative Research in Computer and Communication Engineering, 12(03),* 1710–1715. https://doi.org/10.15680/ijircce.2024.1203056
13. Singh, I., Kumar, N., Srinivasa, K. G., S., Sharma, T., Kumar, V., & Singhal, S. (2020). Database intrusion detection using role and user behavior based risk assessment. *Journal of Information Security and Applications, 55,* 102654. https://doi.org/10.1016/j.jisa.2020.102654

14. Lavrov, E. A., Zolkin, A. L., Aygumov, T. G., Chistyakov, M. S., & Akhmetov, I. V. (2021). Analysis of information security issues in corporate computer networks. *IOP Conference Series: Materials Science and Engineering, 1047(1),* 012117. https://doi.org/10.1088/1757-899x/1047/1/012117
15. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. *In 2008 eighth IEEE international conference on data mining (ICDM).* https://doi.org/10.1109/icdm.2008.17
16. Szumelda, P., Orzechowski, N., Rawski, M., & Janicki, A. (2022). VHS-22 – A very heterogeneous set of network traffic data for threat detection. *In EICC 2022: European interdisciplinary cybersecurity conference. ACM*. https://doi.org/10.1145/3528580.3532843
17. De Carvalho Bertoli, G., Pereira Junior, L. A., Saotome, O., Dos Santos, A. L., Verri, F. A. N., Marcondes, C. A. C., Barbieri, S., Rodrigues, M. S., & Parente De Oliveira, J. M. (2021). An end-to-end framework for machine learning-based network intrusion detection system. *IEEE Access, 9,* 106790–106805. https://doi.org/10.1109/access.2021.3101188
18. Hulak, H. M., Zhiltsov, O. B., Kyrychok, R. V., Korshun, N. V., & Skladannyi, P. M. (2024). *Information and cyber security of the enterprise. Textbook.* Lviv: Publisher Marchenko T. V.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (27), 2025
ISSN 2663 - 4023

Yurii Shchavinskyi
Candidate of Technical Sciences
Associate Professor, Associate Professor of Information and Cyber Security Department
State University of Information and Communication Technologies, Kyiv, Ukraine
ORCID ID: 0000-0002-2319-8983
yushchavinsky@ukr.net

Oleksandr Budzynskyi
Postgraduate Student, Department of Cybersecurity and Information Protection Management
State University of Information and Communication Technologies, Kyiv, Ukraine
ORCID ID: 0009-0002-2402-0711
oleksandr.email@gmail.com

ANALYSIS OF CURRENT PROBLEMS OF SECURITY OF CORPORATE DATABASES IN THE CONDITIONS OF MODERN INFRASTRUCTURE AND WAYS TO SOLUTION THEM

Abstract. This research focuses on analyzing the current security challenges of corporate databases within modern infrastructure, developing a model for detecting anomalous database access activity, and integrating it into the AlienVault SIEM system for automatic threat response. One of the main issues in database security is the need for immediate anomaly detection and response to threats affecting database availability, confidentiality, and integrity. The analysis of scientific literature led to the conclusion that modern infrastructure significantly changes the approach to corporate database security, creating both new opportunities and threats. The challenges associated with the impact of modern infrastructure on database security require new ways to solve problems and a proactive integrated approach, which consists in applying artificial intelligence in organizing the protection of corporate databases.. The study employs the IsolationForest algorithm to develop an anomaly detection model for database access, utilizing open-source Python libraries. The model was trained on historical data, followed by testing and evaluating its effectiveness using Accuracy, Precision, Recall, and ROC AUC metrics. The training process achieved a high level of anomaly detection accuracy (Accuracy = 98.8%, ROC AUC ≈ 0.99, Precision = 0.86, Recall = 0.99). The model's integration into AlienVault was implemented through an external script execution mechanism. The developed model enables real-time identification of potential threats, risk assessment, and automatic blocking of malicious requests or sending alerts to the Security Operations Center (SOC). A novel approach to integrating machine learning algorithms into SIEM systems has been proposed, ensuring proactive anomaly detection and response to database security threats. Implementing this model enhances corporate database security, reduces the risk of data leaks, and ensures prompt responses to cybersecurity incidents.

Keywords: cybersecurity; database security; machine learning; SIEM; AlienVault; IsolationForest; SOC.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Kostiuk, Y., Bebeshko, B., Kriuchkova, L., Lytvynov, V., Oksanych, I., Skladannyi, P., & Khorolska, K. (2024). Information protection and data exchange security in wireless mobile networks with authentication and key exchange protocols. *Cybersecurity: Education, Science, Technique, 1(25),* 229–252. https://doi.org/10.28925/2663-4023.2024.25.229252
2. Kyrychok, R. V., Skladannyi, P. M., Buryachok, V. L., Hulak, G. M., & Kozachok, V. A. (2016). Problems of ensuring control over the security of corporate networks and ways to solve them. *Scientific Notes of the Ukrainian Research Institute of Communications, 3(43),* 48–61.
3. Adenubi, A. O., & P. Oduroye, A. (2024). Data security in big data: challenges, strategies, and future trends. *International journal of research in education humanities and commerce, 05(02),* 01–15. https://doi.org/10.37602/ijrehc.2024.5201

4. Li, X., Wang, Z., Leung, V. C. M., Ji, H., Liu, Y., & Zhang, H. (2021). Blockchain-empowered data-driven networks. *ACM Computing Surveys, 54(3),* 1–38. https://doi.org/10.1145/3446373
5. Bao, R., Chen, Z., & Obaidat, M. S. (2018). Challenges and techniques in Big data security and privacy: A review. *Security and Privacy, 1(4),* Article e13. https://doi.org/10.1002/spy2.13
6. Yeoh, W., Liu, M., Shore, M., & Jiang, F. (2023). Zero trust cybersecurity: Critical success factors and a maturity assessment framework. *Computers & Security, 103412.* https://doi.org/10.1016/j.cose.2023.103412
7. Prince, N. U., Faheem, M. A., Khan, O., Hossain, K., Alkhayyat, A., Hamdache, A., & Elmouki, I. (2024). AI-Powered data-driven cybersecurity techniques: boosting threat identification and reaction. *Nanotechnology Perceptions 20(S10),* 332–353. https://doi.org/10.62441/nano-ntp.v20is10.25
8. Paul, P., & Aithal, P. S. (2019). Database Security: An overview and analysis of current trend. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3497728
9. Yasmeen, M. (2018). NOSQL database engines for big data management. *International Journal of Trend in Scientific Research and Development, 2(6),* 617–622. https://doi.org/10.31142/ijtsrd18608
10. Mousa, A., Karabatak, M., & Mustafa, T. (2020). Database security threats and challenges. *8th international symposium on digital forensics and security (ISDFS).* https://doi.org/10.1109/isdfs49300.2020.9116436
11. Zaw, T. M., Thant, M., & Bezzateev, S. V. (2019). Database security with AES encryption, elliptic curve encryption and signature. *2019 wave electronics and its application in information and telecommunication systems (WECONF).* https://doi.org/10.1109/weconf.2019.8840125
12. Mataz, A., & Haripriya, V. (2024). Enhancing data protection through advanced encryption or improving data security with advanced encryption. *International Journal of Innovative Research in Computer and Communication Engineering, 12(03),* 1710–1715. https://doi.org/10.15680/ijircce.2024.1203056
13. Singh, I., Kumar, N., Srinivasa, K. G., S., Sharma, T., Kumar, V., & Singhal, S. (2020). Database intrusion detection using role and user behavior based risk assessment. *Journal of Information Security and Applications, 55,* 102654. https://doi.org/10.1016/j.jisa.2020.102654
14. Lavrov, E. A., Zolkin, A. L., Aygumov, T. G., Chistyakov, M. S., & Akhmetov, I. V. (2021). Analysis of information security issues in corporate computer networks. *IOP Conference Series: Materials Science and Engineering, 1047(1),* 012117. https://doi.org/10.1088/1757-899x/1047/1/012117
15. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. *In 2008 eighth IEEE international conference on data mining (ICDM).* https://doi.org/10.1109/icdm.2008.17
16. Szumelda, P., Orzechowski, N., Rawski, M., & Janicki, A. (2022). VHS-22 – A very heterogeneous set of network traffic data for threat detection. *In EICC 2022: European interdisciplinary cybersecurity conference. ACM.* https://doi.org/10.1145/3528580.3532843
17. De Carvalho Bertoli, G., Pereira Junior, L. A., Saotome, O., Dos Santos, A. L., Verri, F. A. N., Marcondes, C. A. C., Barbieri, S., Rodrigues, M. S., & Parente De Oliveira, J. M. (2021). An end-to-end framework for machine learning-based network intrusion detection system. *IEEE Access, 9,* 106790–106805. https://doi.org/10.1109/access.2021.3101188
18. Hulak, H. M., Zhiltsov, O. B., Kyrychok, R. V., Korshun, N. V., & Skladannyi, P. M. (2024). *Information and cyber security of the enterprise. Textbook.* Lviv: Publisher Marchenko T. V.