Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка

Факультет інформаційних технологій
Кафедра кібербезпеки та захисту інформації

ПОЯСНЮВАЛЬНА ЗАПИСКА

Дипломної роботи
магістра
(назва освітньо-кваліфікаційного рівня)

галузь знань *12 Інформаційні технології*
(шифр і назва галузі знань)

спеціальність *125 Кібербезпека*
(код і назва спеціальності)

освітній ступень *магістр*
(назва освітньої програми)

освітньо-наукова програма *кібербезпека*

на тему: «*Використання машинного навчання для запобігання витоку в корпоративних мережах*»

Виконавець: студентка ІІ курсу, групи КБм-21

(підпис) *Кравченко Людмила Денисівна*
(прізвище ім'я по-батькові)

	Прізвище, ініціали	Оцінка	Підпис
Науковий керівник	Наконечний В. С.		

Рецензент	Сайко В.Г.		

Нормоконтроль	Даков С.Ю.		

Київ 2022

Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка

Факультет інформаційних технологій
Кафедра кібербезпеки та захисту інформації

ЗАТВЕРДЖЕНО:
завідувач кафедри кібербезпеки
та захисту інформації
_____ Н.В. Лукова-Чуйко
«__» _____ 2021 р.

ЗАВДАННЯ
на виконання дипломної роботи

спеціальності *125 Кібербезпека*
(код і назва спеціальності)

студентки *КБм-21* (група) *Кравченко Людмили Денисівни* (прізвище ім'я по-батькові)

Тема дипломної роботи *Використання машинного навчання для запобігання витоку інформації в корпоративних мережах*

1. ПІДСТАВИ ДЛЯ ПРОВЕДЕННЯ РОБОТИ

Рішення засідання кафедри кібербезпеки та захисту інформації факультету інформаційних технологій протокол № 5 від 29.10.2021

2. МЕТА ТА ВИХІДНІ ДАНІ ДЛЯ ПРОВЕДЕННЯ РОБІТ

Об'єкт досліджень	Процес захисту корпоративних мереж від витоку інформації за допомогою машинного навчання.
Предмет досліджень	Механізми захисту корпоративних мереж за допомогою алгоритмів машинного навчання.
Мета	Розробка рекомендацій щодо застосування алгоритмів машинного навчання для забезпечення інформаційної безпеки корпоративних мереж.
Вихідні дані для проведення роботи	Методи захисту від витоку даних платіжних карток через інтернет-браузер.

3. ОЧІКУВАНІ НАУКОВІ РЕЗУЛЬТАТИ

Наукова новизна вперше запропоновані рекомендації щодо запобігання витоку інформації; порівнянні системи для запобігання витоку інформації в корпоративних мережах.

Практична цінність Полягає у розробці рекомендації щодо вибору алгоритмів ML від витоку даних в корпоративних мережах.

4. ВИМОГИ ДО РЕЗУЛЬТАТІВ ВИКОНАННЯ РОБОТИ

Робота виконана у повному обсязі відповідно до теми.

5. ЕТАПИ ВИКОНАННЯ РОБОТИ

Найменування етапів робіт	Строки виконання робіт (початок-кінець)
Розробка плану для досягнення мети роботи	29.10.2021 – 23.01.2022
Аналіз літературних джерел	24.01.2022 – 14.02.2022
Розробка рекомендацій алгоритмів ML від витоку даних в корпоративних мережах.	15.02.2022 – 24.04.2022
Оформлення і друк пояснювальної записки	25.04.2022 – 19.05.2022

6. РЕАЛІЗАЦІЯ РЕЗУЛЬТАТІВ ТА ЕФЕКТИВНІСТЬ

Економічний ефект Зниження збитків підприємству через викрадення конфіденційних даних.

Соціальний ефект Покращення захисту корпоративних мереж від витоку.

7. ДОДАТКОВІ ВИМОГИ

Завдання видав _____ ___Наконечний В. С.___
(підпис) (прізвище, ініціали)

Завдання прийняв

до виконання  __Кравченко Л.Д.
(підпис) (прізвище, ініціали)

Дата видачі завдання: _____
Термін подання дипломної роботи до ЕК _____

РЕФЕРАТ

Пояснювальна записка до дипломної роботи «Використання машинного навчання (ML) для запобігання витоку інформації в корпоративних мережах»: складається зі вступу, трьох розділів, висновків та списку використаних джерел. Загальний обсяг роботи – 59 сторінок. Робота містить 13 рисунків, 1 таблицю. Список використаних джерел включає 65 джерел.

Об’єкт дослідження — процес захисту корпоративних мереж від витоку інформації за допомогою машинного навчання.

Предмет дослідження — механізми захисту корпоративних мереж за допомогою алгоритмів машинного навчання.

Мета роботи — розробка рекомендацій щодо застосування алгоритмів машинного навчання для забезпечення інформаційної безпеки корпоративних мереж.

Методи дослідження – структурний аналіз та порівняння способів захисту корпоративних мереж.

У роботі проаналізовано проблеми витоку інформації в корпоративних мережах та досліджено типи атак та механізми захисту корпоративних мереж. Проведено аналіз методів машинного навчання та його використання в кібербезпеці. Досліджено шляхи протидії атакам за допомогою ML. Розроблені рекомендації щодо застосування алгоритмів ML від витоку даних.

Наукова новизна: вперше запропоновані рекомендації щодо запобігання витоку інформації; порівнянні системи для запобігання витоку інформації в корпоративних мережах.

Актуальність теми: методи машинного навчання дозволяють замінювати класичні технології, тому що мають перевагу за якістю та швидкістю функціонала. До того ж ML може автоматизувати процеси протидії витокам інформації в компаніях, на які раніше витрачалося багато людських ресурсів. За допомогою

машинного навчання можна швидко обробляти та аналізувати величезні обсяги даних.

Ключові слова: машинне навчання, кібербезпека, корпоративні мережі, кіберзахист з використанням машинного навчання.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

AE – adversarial examples або “суперечний приклад”, дані, які ML модель не може правильно ідентифікувати

DLP – data leak prevention, система захисту від витоку конфіденційних даних

IDS – Intrusion Detection System, система виявлення вторгнень

IPS – Intrusion Prevention System, система запобігання вторгненням

ML – Machine Learning, машинне навчання

NBAD – Network Behavior and Anomaly Detection, поведінка мережі та виявлення аномалій

RDP – Remote Desktop Protocol, протокол віддаленого робочого столу.

SIEM – security information and event management, управління інформаційною безпекою та подіями безпеки

UEBA – user and entity behavior analytics, аналіз поведінки користувачів та сутностей

VML – vector machine learning, вектор машинного навчання

VPN – virtual private network, віртуальна приватна мережа

АС – автоматизована система

ІБ – інформаційна безпека

ІТ – інформаційні технології

ЗІ – захист інформації

НД ТЗІ – нормативний документ технічного захисту інформації

ОС – обчислювальна система

ПБ – політика безпеки

ПЗ – програмне забезпечення

ПК – персональний комп’ютер

ШІ – штучний інтелект

ЗМІСТ

ВСТУП

Завдяки прискореній цифровізації, яку багато компаній зазнали протягом останніх років, можна спостерігати також збільшення численних кібератак [1]. І кожного року це питання буде ставати найбільш гострим та актуальним.

Найбільше проблема виникає, коли користувач припускає, що, маючи правильні облікові дані для доступу, він може вільно переміщатися по системі та сподіватися, що нічого не станеться. Але, на жаль, незалежно від того, скільки способів захисту інформації (ЗІ) існує в мережі, щойно користувач натискає на шкідливе посилання або вкладення, це дозволяє кіберзлочинцям скомпрометувати системи. Саме тому фішинг і програми-вимагачі завдають такої шкоди корпоративним мережам.

Близько 281,5 мільйона людей постраждали від зламу даних у 2021 році [2]. Тим часом кіберзлочинність коштує компаніям 1,79 мільйона доларів за хвилину, залишаючи перед нами досить тривожну картину кібербезпеки. Більшість топменеджерів бачать кібербезпеку як оперативний пріоритет для захисту свого бізнесу [3].

Для запобігання витоку інформації в корпоративних мережах спеціалісти все частіше починають застосовувати технології машинного навчання (ML). Ці методи дозволяють замінювати класичні технології, тому що мають перевагу за якістю та швидкістю функціонала. До того ж ML може автоматизувати процеси в компаніях, на які раніше витрачалося багато людських ресурсів. За допомогою ML можна швидко обробляти та аналізувати величезні обсяги даних.

Значний вклад в розвиток машинного навчання в Україні внесли В. Радченко, Є. Терпіль, О. Петрів, А. Чернятевич, О. В. Антонюк , Н.Ф. Тамайо, М. М. Сажок, О. П. Ігнатенко, В. Тимчишин, А. О. Фролова, Т. Шевченко, В. Семенов, Т. Кучеренко, О. Жиліч та інші.

Серед закордонних науковців варто згадати Д. Хемпель, E. Kouskoumvekaki, N. Nikolaidis, R. Ohbuchi, I. Pitas, V. Solachidis, M. Voigt та інших.

Об'єкт дослідження: процес захисту корпоративних мереж від витоку інформації за допомогою машинного навчання.

Предмет дослідження: механізми захисту корпоративних мереж за допомогою алгоритмів ML.

Мета роботи — розробка рекомендацій щодо застосування алгоритмів машинного навчання для забезпечення інформаційної безпеки (ІБ) корпоративних мереж. Для її досягнення було визначено такі завдання:

- Провести аналіз існуючих проблем витоку інформації в корпоративних мережах.
- Дослідити типи атак та механізми захисту корпоративних мереж.
- Порівняти засоби захисту корпоративних мереж.
- Зробити аналіз методів машинного навчання.
- Дослідити та проаналізувати способи використання машинного навчання в кібербезпеці.
- Розглянути способи використання ML для захисту інформації в корпоративних мережах.
- Дослідити шляхи протидії атакам за допомогою ML.
- На основі проведеного аналізу розробити рекомендації щодо вибору алгоритмів ML від витоку даних

Результати даного дослідження можуть бути корисними для фахівців із кібербезпеки, які займаються захистом корпоративних мереж.

Наукова новизна дослідження полягає в тому, що вперше порівняні системи для запобігання витоку інформації в корпоративних мережах та запропоновані рекомендації щодо вибору алгоритмів машинного навчання.

Апробація результатів роботи: основні наукові положення і результати роботи апробовані на V Міжнародній науково-практичній конференції “Проблеми кібербезпеки інформаційно-телекомунікаційних систем” (PCSITS)” (Київ, 2022). Основні положення дипломної роботи викладені у матеріалах зазначеної наукової конференції.

РОЗДІЛ 1
ПРОБЛЕМИ ВИТОКУ ІНФОРМАЦІЇ В КОРПОРАТИВНИХ МЕРЕЖАХ

1.1 Загрози інформаційній безпеці корпоративним мережам організації

Необхідність захистити свій бізнес від кібератак ще ніколи не була такою важливою, як сьогодні. Адже кількість ризиків постійно зростає [1, 4]. Незалежно від того, зберігаються дані та інформація компанії на жорсткому диску чи надсилаються електронною поштою, потрібно остерігатися витоку даних.

Метою порушників є передача інформації третім особам за межі автоматизованих систем (АС) з метою подальшого несанкціонованого використання. Хакерські атаки загрожують невеликим компаніям так само як і гігантським організаціям. Навіть окремі операції з присутністю на вебсайті або інтернет-магазині можуть бути виявлені хакерами, які шукають конфіденційні дані про клієнтів. Якщо у мережі є незначне, на вигляд посилання, існує висока ймовірність того, що хтось знайде його і використає на свою користь.

Організаціям різних сфер діяльності доводиться працювати в умовах високої складності, через динамічність та невизначеність сьогодення. Поширеність засобів захисту інформації знімає тільки одну частину проблеми – випадкові витоки, але ніяк не впливає на зловмисні. Використання SIEM та DLP-систем перекривають лише часткові витоки інформації [5].

У світі, де кіберзлочинці продовжують щодня винаходити нові зловмисні методи атак, захист корпоративних мереж потребує комплексний підхід з урахуванням нових технологій машинного навчання.

1.1.1 Класифікація загроз безпеці комп'ютерним системам та мережам організації

Під загрозою кібербезпеки розуміють будь-яку можливу зловмисну атаку, яка спрямована на незаконний доступ до даних, порушення цифрових операцій або пошкодження інформації [6]. Кіберзагрози можуть походити від різних дійових осіб, включаючи корпоративних шпигунів, активістів, терористичних груп, ворожих національних держав, злочинних організацій, хакерів-одинаків та незадоволених співробітників.

Кіберзловмисники можуть використовувати конфіденційні дані окремої особи або компанії, щоб викрасти інформацію або отримати доступ до їхніх фінансових рахунків, серед інших потенційно шкідливих дій. Типи загроз ІБ дуже різноманітні та мають безліч класифікацій [7].

1. За аспектом ІБ, на які направлені загрози:

- Загроза конфіденційності — коли особиста інформація розкривається третій стороні без згоди власника.
- Загроза цілісності — пов'язані із вірогідним зміненням або спотворенням інформації.
- Загроза доступності — пов'язані з неможливістю доступу до важливих даних.

2. За розташуванням джерела загроз:

- Внутрішні (розташовані всередині системи).
- Зовнішні (перебувають поза системою).

3. За розмірами завданого збитку загрози класифікують:

- Загальні — заподіяння значної шкоди [8].
- Приватні — заподіяння шкоди окремим властивостям елементів об'єкта безпеки [9].
- Локальні — заподіяння шкоди окремим частинам об'єкта безпеки.

4. За ступенем впливу на інформаційну систему:

- Активні — структура і зміст системи піддається змінам.

- Пасивні — структура і зміст системи не змінюються.

5. За природою виникнення:

- Природні — внаслідок дії природних явищ або стихійних лих.
- Штучні — через вплив суб'єктів на мережу. Штучні загрози своєю чергою поділяють на:

- Ненавмисні загрози: такі як помилки співробітників, збої обчислювальної техніки, відмови в роботі систем;
- Навмисні загрози: несанкціонований доступ суб'єктів до інформації, встановлення шкідливого програмного забезпечення (ПЗ), розповсюдження вірусних програм тощо.

Згідно зі статистикою від фахівців ІБ понад 87% організацій зіштовхнулися зі спробою використання відомих вже вразливостей [10].

Крім того, зловмисники, мотивовані фінансовими міркуваннями, продовжують проводити шкідливі кампанії. Шахраї розвивають методи, щоб використовувати голосовий фішинг (вішинг), подвійне застосування програм-вимагачів, перехоплення потоків електронної пошти та атаки, націлені на хмарні інфраструктури.

Такі результати досліджень є підставою для акценту на ефективному впровадженні комплексних систем захисту інформації [11].

1.1.2 Джерела та канали витоку конфіденційної інформації

Економічна безпека будь-якої організації напряму залежить від інформаційної складової. Через це питання ІБ повинно бути пріоритетом для будь-яких організацій.

Сьогодні інсайдерська інформація може бути дуже небезпечною. Завдання керівників служби підтримати безпеку даних та протистояти загрозам за допомогою впровадження ефективних систем комплексного захисту [12, 13].

Нижче наведено п’ять найпоширеніших загроз безпеки корпоративної мережі [13]:

1. Фішинг

Цей тип шахрайства в Інтернеті призначений для крадіжки конфіденційної інформації, такої як номери кредитних карток та паролі. Фішингові атаки видають себе за авторитетні банківські установи, вебсайти та особисті контакти, які надходять у вигляді негайних фішингових електронних листів або повідомлень, створених таким чином, щоб виглядати законними.

Після натискання URL-адреси або відповіді на повідомлення вам буде запропоновано ввести фінансові дані або використати облікові дані, які потім надсилають дані до шкідливого джерела.

2. Комп’ютерні віруси

Це частини програмного забезпечення, призначені для поширення з одного комп’ютерного пристрою на інший. Здебільшого віруси завантажуються з певних вебсайтів або надсилаються як вкладення електронної пошти з метою заразити комп’ютер, а також інші комп’ютери у списку контактів через системи у мережі. Вони можуть вимкнути налаштування безпеки, розсилати спам, красти та пошкоджені дані з комп’ютера і навіть видаляти всі файли на жорсткому диску.

3. Шкідливе програмне забезпечення

Шкідливе ПЗ в основному використовується злочинцями для утримання системи, крадіжки конфіденційних даних або встановлення шкідливих програм на пристрої без відома власника. Він поширює шпигунські програми, троянські програми через спливаючі вікна, заражені файли, фіктивні вебсайти або повідомлення електронної пошти. Це може перешкоджати запуску програм, шифруванню файлів і навіть повному використанню пристрою.

4. Зловмисне програмне забезпечення безпеки

Це шкідливе ПЗ, яке обманює користувачів, змушуючи їх повірити, що їхні заходи безпеки не оновлені, або на комп’ютері є вірус. Потім вони пропонують допомогти вам встановити або оновити налаштування безпеки користувача, попросивши вас заплатити за інструмент або завантажити їхню програму, щоб

допомогти усунути ймовірні віруси. Це може призвести до встановлення фактичного шкідливого ПЗ на пристрої.

5. Атака відмови в обслуговуванні

Відмова в обслуговуванні намагається перешкодити легальним користувачам отримати доступ до послуг або інформації з вебсайту [14]. Це трапляється, коли зловмисники перевантажують вебсайт трафіком. Це виконується одним комп'ютером і його підключенням до Інтернету, що може дозволити зловмиснику отримати доступ до облікових даних. Розподілена відмова в обслуговуванні подібна до відмови в обслуговуванні, але її важче подолати. Це тому, що цей тип атаки запускається з різних комп'ютерів, які поширюються по всьому світу. Мережа з цих зламаних комп'ютерів називається ботнетом.

Основними проблемами та загрозами для мережевої безпеки станом на 2020-2021 рік були [15]:

1. Атаки на ланцюги постачання

8 грудня 2020 року компанія з кібербезпеки FireEye повідомила, що виявила шкідливе програмне забезпечення Sunburst у власних мережах [16]. Розслідування цієї інфекції виявило масштабну кампанію кібератак, яка вплинула на 18 000 організацій, 425 компаній зі списку Fortune 500 (включаючи Microsoft), а також на державні установи. Шкідливе програмне забезпечення SUNBURST розповсюджувалося через скомпрометовані оновлення ПЗ для керування мережею SolarWinds Orion . Зловмисникам вдалося скомпрометувати SolarWinds за допомогою нової атаки на її облікові записи Office 365, що дозволило їм підробити маркер Azure Active Directory для привілейованого облікового запису та використовувати зламані облікові дані адміністратора, щоб отримати доступ до сервера керування оновленнями компанії [17].

Під час розробки зловмисники змогли модифікувати оновлення, щоб включити зловмисне програмне забезпечення. Такий широкий діапазон атаки зробив це найуспішнішою відомою атакою на сьогодні.

Щоб запобігти майбутнім атакам, потрібно застосувати найкращі методи безпеки, такі як [17-18]:

- Найменші привілеї та сегментація мережі: ці найкращі методи можуть допомогти відстежувати та контролювати переміщення в мережі організації.
- DevSecOps: Інтеграція безпеки в життєвий цикл розробки може допомогти виявити, чи програмне забезпечення (наприклад, оновлення Orion) було зловмисно змінено.
- Автоматизоване запобігання загрозам і пошук загроз: аналітики центрів безпеки (SOC) повинні активно захищатися від атак у всіх середовищах, включаючи мережу, кінцеву точку, хмару та мобільний пристрій.

2. Вішинг

Хоча фішинг є найвідомішим типом атаки соціальної інженерії, інші методи можуть бути настільки ж ефективними [18]. Через телефон зловмисник може використовувати методи соціальної інженерії, щоб отримати доступ до облікових даних та іншої ключової інформації або переконати жертву відкрити файл або встановити шкідливе програмне забезпечення. Вішинг є перспективною загрозою корпоративній кібербезпеці. Загроза вішингу лише посилюватиметься, оскільки технологія запису глибоких фейків удосконалюється та стає все більш доступною[19-20].

Вішинг — це низькотехнологічна атака, а це означає, що освіта співробітників є важливою для захисту від неї. Компанії можуть навчити співробітників не відмовлятися від конфіденційної інформації та самостійно перевіряти ідентифікацію абонента, перш ніж виконувати запити.

3. Програми-вимагачі

Програми-вимагачі були однією з найдорожчих кіберзагроз для організацій у 2020 році [21]. У 2020 році це обійшлося компаніям у 20 мільярдів доларів проти 11,5 мільярдів доларів у 2019 році. У третьому кварталі 2020 року середній платіж із викупом становив 233 817 доларів США, що на 30% більше, ніж у попередньому кварталі [22].

Зростання цих атак подвійного вимагання означає, що організації повинні прийняти стратегію запобігання загрозам, а не покладатися лише на виявлення чи усунення. Стратегія, спрямована на профілактику, повинна включати:

- Рішення для боротьби з програмним забезпеченням-вимагачем. Організації повинні застосовувати рішення безпеки, розроблені спеціально для виявлення та ліквідації зараження програм-вимагачів у системі.
- Управління вразливістю. Виправлення вразливих систем або використання технологій віртуального виправлення, таких як система запобігання вторгненням (IPS), необхідно для закриття поширених векторів зараження програм-вимагачів, таких як протокол віддаленого робочого столу (RDP).
- Навчання співробітників: детальне поінформування про ризики відкриття вкладень у шкідливих електронних листах або натискання на них посилань.

4. Уразливості віддаленого доступу

Сплеск віддаленої роботи через COVID-19 зробив віддалений доступ загальною метою кіберзлочинців у 2020 році [23]. У першій половині року різко зросла кількість атак на технології віддаленого доступу, такі як RDP і VPN. Щодня було виявлено майже мільйон атак на RDP. Кіберзлочинці зосередилися на вразливих VPN-порталах, шлюзах і додатках, оскільки стало відомо про нові вразливості в цих системах. У мережі датчиків Check Point зросла кількість атак на вісім відомих уразливостей у пристроях віддаленого доступу, включаючи Cisco і Citrix [24].

Щоб керувати ризиками вразливостей віддаленого доступу, організації повинні виправляти вразливі системи безпосередньо або розгортати технології віртуального виправлення, такі як IPS. Вони також повинні захищати віддалених користувачів шляхом розгортання комплексного захисту кінцевих точок із технологіями виявлення та реагування кінцевих точок (EDR) для покращення виправлення та пошуку загроз.

5. Мобільні загрози

Завдяки віддаленій роботі використання мобільних пристроїв різко зросло. На жаль, мобільні пристрої також були мішенню для великих кампаній шкідливого ПЗ, включаючи банківське зловмисне програмне забезпечення.

Підприємства можуть захистити мобільні пристрої співробітників за допомогою легкого мобільного рішення безпеки для некерованих пристроїв. Вони також повинні навчити користувачів захищатися, встановлюючи лише програми з офіційних магазинів додатків, щоб мінімізувати ризик.

Сьогодні на ринку існує досить багато рішень, що дозволяють визначати та запобігати витоку конфіденційної інформації з тих чи інших каналах — IDS, IPS, DLP, SIEM, NBAD системи. Підприємствам потрібна цілісна система у загальнодоступних хмарних середовищах та розгортання уніфікованих, автоматизованих засобів захисту, що покриває всі види комунікації, здатний ефективно захищати дані. Це дозволятиме попри різнорідний індекс різних каналів витоку, забезпечити цілісність конфіденційність та доступність до інформаційних активів. Саме принцип комплексності узятий за основу при розгляді рішень для боротьби з витоками даних [25].

1.2 Нормативно-правова основа захисту інформації

У даній роботі враховувались такі нормативні документи у сфері ЗІ:

1. Закон України Про Інформацію

“Даний закон визначає основні принципи інформаційних відносин та державної інформаційної політики; суб’єктів та об’єктів інформаційної політики; право та гарантії права на інформацію; основні види інформаційної діяльності; відповідальність за порушення законодавства про інформацію” [26].

2. НД ТЗІ 1.1-002-99 «Загальні положення щодо захисту інформації в комп'ютерних системах від несанкціонованого доступу»

Згідно з загальними положеннями цього документу [27]: Цей нормативний документ технічного захисту інформації (НД ТЗІ) визначає методологічні основи (концепцію) вирішення завдань захисту інформації в комп'ютерних системах і створення нормативних та методологічних документів, що регламентують питання:

- визначення вимог щодо захисту комп'ютерних систем від несанкціонованого доступу;

- створення захищених комп'ютерних систем і засобів їх захисту від несанкціонованого доступу;
- оцінки захищеності комп'ютерних систем і їх придатності для вирішення завдань споживача.

Документ призначено для розробників, споживачів комп'ютерних систем, які використовуються для обробки, збирання, зберігання та передачі конфіденційної інформації.

3. ISO/IEC 15408 «Common Criteria for Information Technology Security Evaluation»

Загальні критерії оцінки безпеки інформаційних технологій (Common Criteria або CC) — це міжнародний стандарт (ISO / IEC 15408) для сертифікації безпеки IT-продуктів. Це структура, яка забезпечує критерії для незалежних, масштабованих і визнаних у всьому світі перевірок безпеки IT-продуктів. Цей стандарт спеціально розроблений для продуктів, призначених для ринків із високим рівнем безпеки, таких як державний, банківський або військовий сектори. Тому сертифікація відповідно до цього стандарту є важливою. Стандарт містить два основних види вимог безпеки: функціональні та вимоги довіри.

4. ISO 27001 — серія стандартів, яка забезпечує інформаційну безпеку організацій. Стандарт ISO/IEC 27001 широко відомий, що передбачає вимоги до системи управління інформаційною безпекою (ISMS), хоча в сімействі ISO/IEC 27000 існує більше десятка стандартів. Їх використання дозволяє організаціям будь-якого виду керувати безпекою активів, таких як фінансова інформація, інтелектуальна власність, дані співробітників або інформація, довірена третіми сторонами.

1.3 Порівняння засобів захисту корпоративних мереж

Засобів захисту корпоративних мереж існує багато [27-28], та чи справді вони актуальні досі розглянемо нижче. За традиційного підходу, коли використовуються фаєрволи, антивіруси, DLP та інші засоби захисту, вони орієнтуються на боротьбу з

чимось наперед відомим. Натомість ML працює за іншим принципом — заздалегідь навчає систему, після чого вона починає видавати нам вердикт за новими, раніше не відомими даними: погана це або хороша поведінка, файл, запит в інтернет, активність користувача, програми тощо.

Проведемо аналіз основних систем захисту корпоративних мереж:

- IDS: Система виявлення вторгнень — це пасивний моніторинг, який може виявити несанкціоновану діяльність.

Усе, що виявлено IDS, повідомляється назад до головного місця запису журналу; є два типи - NIDS і HIDS. NIDS — це мережевий IDS, який зазвичай виконується шляхом активації моніторингу на порту або встановлення мережного крана на лінії. HIDS — це IDS на основі хоста, який встановлюється та запускається на цільовій комп’ютерній системі; на кожному комп’ютері в корпоративному середовищі має працювати якийсь пакет HIDS.

- IPS: Система запобігання вторгненням — це активна система моніторингу, яка виявляє і реагує на спроби несанкціонованого доступу. Ця система може блокувати ці спроби, закриваючи порти або блокуючи шляхи або обмежуючи рівні доступу.

- DID: Defense In Depth — це застосування IDS та IPS одночасно у мережі підприємства для виявлення та запобігання несанкціонованому доступу або контролю чутливого обладнання та даних. Складні евристики використовують алгоритми для визначення автентичності трафіку та облікових даних для кожного запиту.

- SIEM: Безпечне управління інформацією та подіями — це програми та ПЗ, які адміністратори та менеджери використовують для захисту свого середовища. Існує кілька різних компаній із програмним забезпеченням SIEM, а також хмарними пропозиціями SIEM як сервіс.

- DLP: Запобігання втраті даних — це частина теорії, частина політики та частина ПЗ. Запобігання втраті даних – це директива, доповнена програмним забезпеченням, запроваджена політикою та застосовна до кожної людини, яка взаємодіє з інформаційним середовищем.

● NBAD: Мережеву поведінка та виявлення аномалій — це набір інструментів, які працюють у поєднанні з іншими інструментами виявлення та запобігання для моніторингу, виявлення, звітування та реагування на спроби несанкціонованого доступу для захисту інформації або систем.

У таблиці 1.1 наведене порівняння підходів до захисту корпоративних мереж за допомогою машинного навчання та різноманітних систем запобігання витоку .

Таблиця 1.1

Порівняльна характеристика підходів до захисту інформаційної безпеки

Погрози ІБ	Старий підхід	Новий підхід з ML
Шкідливий код	Антивіруси, IDS тощо. використовують сигнатурні методи для детектування шкідливого коду Бореться з відомою вразливістю	Розпізнавання образів (за схожістю) та предикативна аналітика
DDoS-атаки	Аналітики спостерігають за мережевим трафіком для виявлення DDoS атак Вимагає ресурси, залежить від людини	Алгоритми автоматично детектують ненормальну активність Швидка реакція, мінімум ресурсів
Фішингові домени	Аналітики фіксують домени, з яких реалізуються атаки та поміщають їх у чорні списки	Аналіз взаємозв'язків та передбачення DGA Запобігання
Соціальний інжиніринг	Вивчення тактики хакерів Схильне до помилок	Навчання + контроль Поведінка + біометрія Зниження числа помилок

Виходячи з результатів порівняння, наведених у таблиці, можемо зробити висновок, що застосування виключно IDS, IPS, DLP, SIEM систем без використання методів ML не є ефективним. Через те, що такі системи захищають дані від вже наявних загроз.

1.4 Постановка задачі з захисту мережі організації від витоку інформації.

Управління організаціями вимагає одночасно стабільного функціонування всіх систем, для чого необхідно багато ресурсів різного роду. Наприклад, обробки та зберігання великої кількості інформації, яка циркулює у корпоративній мережі. Потрібно пам'ятати та враховувати при цьому доступи співробітників до цього обсягу даних та засобів ІТС (електронна пошта, віддалене управління ПК, віддалений доступ до внутрішніх локальних мереж підприємства, внутрішні сервери обміну та зберігання даних). Якщо інформація про проєктні угоди або інформацію про тендер витікає, ваш бізнес може втратити значні доходи та репутацію.

Витоки інформації не завжди можуть заважати вашому бізнесу проте часто виникають непрямі наслідки. Витік конфіденційної інформації про клієнтів може зашкодити репутації вашої компанії на ринку. Майбутні клієнти будуть боятися працювати з вами або розкривати особисту інформацію вашій компанії.

Загрози інформаційній безпеці мають комплексний характер: зовнішні зловмисники здійснюють атаки на мережі та інформаційні ресурси організацій, а власні співробітники часто стають джерелами конфіденційної інформації для третіх осіб.

Зовнішні організації (конкуренти, преса, наглядові органи) і, на жаль, співробітники, що мають легальний доступ до оброблюваної інформації, зацікавлені в діставанні доступу до багатьох категорій оброблюваної інформації, зокрема до відомостей про клієнтів і історію роботи з ними, персональних даних співробітників, документів стратегічного розвитку, внутрішніх аналітичних звітів і багато чому іншому.

Висновок до першого розділу

Керівники безпеки та управління ризиками стикаються з критичним моментом, оскільки цифровий слід організацій розширюється, а централізований контроль кібербезпеки застаріває. Гібридна робота та цифрові бізнес-процеси у хмарі привнесли нові ризики. Водночас складні програми-вимагачі, атаки на

цифровий ланцюжок поставок і вразливості, що глибоко вкоренилися, виявили технологічні прогалини та брак навичок [28-29].

Стає очевидним що запобігання витоку інформації з мережі підприємства стає однією з найважливіших завдань інформаційної безпеки. Використання лише традиційних систем із забезпечення безпеки, таких як IDS, IPS, DLP, SIEM, NBAD, на сьогодні недостатньо.

Сьогодні неможливо розгорнути ефективні технології кібербезпеки, не покладаючись на машинне навчання. Водночас неможливо ефективно розгорнути машинне навчання без всебічного, багатого та повного підходу до базових даних.

Отже, для забезпечення виконання зазначеного завдання необхідно виконати наступні етапи:

- Провести аналіз методів машинного навчання.
- Дослідити та проаналізувати використання машинного навчання в кібербезпеці.
- Розглянути способи використання ML для захисту інформації в корпоративних мережах.
- Дослідити шляхи протидії атакам за допомогою ML.
- На основі проведеного аналізу розробити рекомендації щодо вибору алгоритмів ML від витоку даних.

РОЗДІЛ 2
ЗАСТОСУВАННЯ МАШИННОГО НАВЧАННЯ В КІБЕРБЕЗПЕЦІ

2.1 Методи машинного навчання

Популярність машинного навчання пояснюється такими фактами, як зростання обсягів та різноманітності доступних даних, дешевша й потужніша обчислювальна обробка та доступне зберігання даних. Усе це означає, що можна швидко й автоматично створювати моделі, які можуть аналізувати більші, складніші дані та надавати швидші та точніші результати – навіть у дуже великих масштабах. Створюючи точні моделі, організація має більше шансів визначити прибуткові можливості або уникнути невідомих ризиків [30].

Вбудовані механізми збору інформації про мережу полегшують накопичення необхідних даних для їх подальшого використання в завданнях машинного навчання без необхідності додаткового проміжного програмного забезпечення для цієї мети та з можливістю зміни поведінки мережі на основі отриманих результатів. Завдяки програмовності SDN, максимально ефективні та оптимальні рішення, отримані за допомогою ML, можуть застосовуватися в режимі реального часу [31].

Поміж цього машинне навчання — це технологія, яка допомагає підприємствам ефективно отримувати інформацію з необроблених даних. Зокрема алгоритми машинного навчання можна використовувати для ітераційного навчання з заданого набору даних, розуміння шаблонів, поведінки тощо. Існує велика кількість завдань, у яких можуть бути застосовані алгоритми ML, наприклад, такі як забезпечення безпеки (контроль доступу, виявлення DDOS- та DOS-атак) або прогнозування навантажень та тенденцій у мережі (наприклад, обсягів трафіку).

Цей ітеративний і постійно розвивається характер процесу машинного навчання допомагає підприємствам завжди бути в курсі потреб бізнесу та споживачів. Крім того, простіше, ніж будь-коли, створювати чи інтегрувати ML в наявні бізнес-процеси, оскільки всі основні постачальники хмарних послуг

пропонують платформи машинного навчання. Машинне навчання швидко стає повсюдним у всіх галузях, від сільського господарства до медичних досліджень, фондового ринку, моніторингу трафіку тощо.

2.1.1 Головні визначення

Машинне навчання — це метод аналізу даних, який автоматизує побудову аналітичної моделі. Це розділ штучного інтелекту, заснований на ідеї, що системи можуть вчитися на даних, визначати закономірності та приймати рішення з мінімальним втручанням людини [32]. Все машинне навчання вважається штучним інтелектом (ШІ), але весь ШІ вважається машинним навчанням. Вважається одним з найкращих інструментів штучного інтелекту, який підходить для бізнесу.

ML має три підкатегорії: навчання під наглядом, без нагляду навчання та підкріплення навчання [33]. Після того, як модель виконає самонавчання, вона може почати прогнозувати або приймати рішення щодо нових даних чи ситуацій, які їй передаються. При навчанні без нагляду нема потреби в такому позначеному наборі даних. Однак такий тип навчання не може нічого передбачити. Коли додаються нові дані, модель призначає їх одному з наявних кластерів або створює новий. Підкріплення навчання — це здатність системи взаємодіяти з навколишнім середовищем та визначати найкращі результати [33, 34]. Система «винагороджується» або «карається» за правильну або неправильну відповідь, а на основі отриманих позитивних балів модель готується прогнозувати нові дані.

Модель отримує вхідні дані та робить результат. Весь процес виглядає зображено на рисунку 2.1 [34].

Рисунок 2.1 — Розроблення рішень за допомогою ML.

Головна відмінність між традиційним програмуванням та машинним навчанням у тому, що у машинному навчанні не потрібно будувати модель самостійно. Це завдання виконують алгоритми машинного навчання, хіба що невеликими правками, які дата інженер вносить у налаштування алгоритму.

Переважна більшість завдань, розв'язуваних з допомогою методів машинного навчання, належить до різних видів: навчання з учителем (supervised learning) чи без нього (unsupervised learning). Однак цим учителем зовсім не обов'язково є сам програміст, який стоїть над комп'ютером та контролює кожну дію у програмі. «Вчитель» у термінах машинного навчання – це саме втручання людини у процес обробки інформації. В обох видах навчання машині надаються вихідні дані, які вона має проаналізувати та знайти закономірності. Відмінність лише тому, що навчання з учителем є низка гіпотез, які потрібно спростувати чи підтвердити. Цю різницю легко зрозуміти на прикладах.

Крім названих, розробляються та інші методи навчання: активне, багатозадачне, різноманітне, трансферне і т.д. Особливо успішно розвивається в останні роки «глибоке навчання», при використанні якого можуть успішно поєднуватися алгоритми навчання з вчителем і без вчителя.

2.1.2 Контрольоване навчання

Навчання з учителем — найбільш розвинений і популярний напрямок машинного навчання. Основна ідея полягає в тому, що потрібно задати набір вхідних параметрів та очікуваний результат. Таким чином, навчаєте алгоритм правильним відповідям — звідси й назва.

Для навчання з учителем дані мають бути марковані (labeled). Це означає, що поряд з вхідними параметрами дані повинні містити відповіді або, як заведено називати, ярлики (labels). Наприклад, для завдання прогнозування курсу валют ярликом буде значення курсу обміну валют.

Простіше кажучи, у навчанні з учителем для створення моделі потрібно ставити запитання та надавати відповіді. Приклад моделі зображений на рисунку 2.2.

Рисунок 2.2 — Побудова моделі за допомогою машинного навчання з учителем

Після того, як модель побудована, можемо вимагати відповіді на нові питання. Навчання з учителем може вирішувати два типи завдань: класифікація та регресійний аналіз.

Метод навчання застосовується у випадках, коли є великі обсяги даних. Наприклад тисячі фотографій тварин з маркерами (ярликами): це птах, а це собака. Або ж у медичній діагностиці класифікація визначає наявність чи відсутність у пацієнта певного захворювання. Необхідно створити алгоритм, за допомогою якого машина могла б по фотографії, яку «не бачила» раніше, визначити, хто на ній зображений [35]. У ролі «вчителя» в такому випадку виступає людина, яка

заздалегідь проставила маркери. Машина сама вибирає ознаки, за якими вона відрізняє тварин. Тому надалі знайдений нею алгоритм може бути швидко переналаштований на розв'язання іншої задачі, наприклад, на розпізнавання дітей та птахів. Машина знову-таки сама виконає складну і копітку роботу по виділенню ознак, за якими буде розрізняти об'єкти. А нейромережа, яку навчили розпізнавати котів, може швидко навчити обробляти результати комп'ютерної томографії.

Рисунок 2.3 — Використання моделі класифікатора для визначення об'єкта

Базовий алгоритм аналогічний. Нам потрібний набір зображень, текстів або даних та набір правильних відповідей для кожного з них. Алгоритм машинного навчання отримує ці питання з відповідями та будує модель. Надалі навчена модель може самостійно робити класифікацію нових даних.

Недоліком класифікаційного алгоритму є те, що він може давати відповіді лише на ті питання, яким навчилися. Наприклад, якщо системі задали безліч зображень з собаками та промаркували їх як такі, кінцева модель буде здатна визначати собак на нових зображеннях. Але визначити кота або птаху вона не зможе. Тобто ці алгоритми не підходять для випадків, коли результатом має бути число, яке ML намагається передбачити [36].

Класифікаційні алгоритми працюють тільки для тих випадків, де існує обмежений набір можливих результатів.

Для вирішення завдань з необмеженим набором можливих відповідей є алгоритми регресійного аналізу. Для реалізації такого алгоритму дата інженер слідує вищеописаному процесу. Необхідно зібрати дані, які містять вхідні параметри та правильні відповіді. Ці дані завантажуються в алгоритм регресійного

аналізу, і створює навчену модель. Отримавши модель можемо використовувати її для прогнозування нових значень, використовуючи нові вхідні параметри.

Загалом алгоритми класифікації та регресійного аналізу дуже схожі та відрізняються лише потенційними результатами, які вони можуть зробити.

2.1.3 Неконтрольоване навчання

Хоча маркованих, розмічених даних накопичилося вже досить багато, даних без маркерів (міток) все ж набагато більше. Це зображення без підписів, аудіозаписи без коментарів, тексти без анотацій. Завдання машини при неконтрольованому навчанні — знайти зв'язок між окремими даними, виявити закономірності, підібрати шаблони, упорядкувати дані або описати їх структуру, виконати класифікацію даних [37].

Машинне навчання без вчителя намагається знайти відповіді у немаркованих даних (unlabeled data). Іншими словами, надаємо деякі дані, але не ставимо правильних відповідей. Тому цей тип називається «без вчителя» — алгоритм має самостійно з'ясувати будь-що, без попереднього навчання (рис.2.4).

Рисунок 2.4 — Принцип роботи алгоритму навчання без учителя

Так, машинне навчання без учителя не навчає модель. Натомість використовуються безпосередньо вхідні параметри.

У машинному навчанні без вчителя виділяють три категорії алгоритмів [37]:

1. асоціативні;
2. кластеризація;

3. зниження розмірності.

1. Асоціативні алгоритми

Алгоритм “apriori” дуже популярне рішення для асоціативних завдань. Він дозволяє знаходити предмети чи поняття, які найчастіше використовуються разом. Таким чином, стандартний функціонал типу «покупці, які придбали це, також придбали ось це», може бути реалізований за допомогою певної варіації такого алгоритму. Приклад наведений на рисунку 2.5.

Рисунок 2.5 — Принцип роботи асоціативного алгоритму

У цілому нам необхідно розсортувати інформацію про продукти в різних кошиках і алгоритм виявить комбінації продуктів, що найчастіше зустрічаються.

2. Алгоритми кластеризації

Алгоритми кластеризації дозволяють групувати дані кластери. Один із найпопулярніших алгоритмів у даній категорії – це метод k-середніх (K-Means). На рисунку 2.6 показано, як він працює:

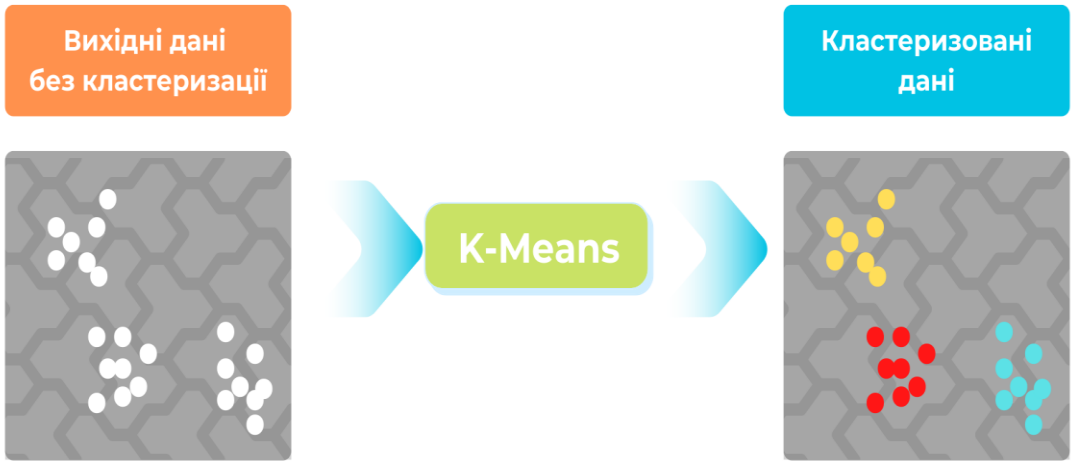

Рисунок 2.6 — Принцип роботи алгоритму кластеризації

Алгоритм працює на один прохід. Нам потрібно внести вихідні дані та алгоритм їх згрупує.

У кластеризації багато сфер застосування [38]:

- групування подібних статей у Google News;
- сегментування ринку для таргетування різних груп покупців;
- об'єднання будинків у райони;
- аналіз соціальних графів визначення груп друзів (в соцмережах);
- кластеризація фільмів за набором властивостей.

3. Зниження розмірності: PCA

У деяких комплексних завданнях машинного навчання сотні чи тисячі вхідних параметрів – це звичайна справа. Обробка такого обсягу даних перевантажує процесор. Чи можна зменшити обсяг вхідних даних без суттєвої втрати інформації?

З цим завданням справляється метод головних компонентів (Principal Component Analysis, або PCA). Концепція функціонування PCA зображена на рисунку 2.7.

Рисунок 2.7 — Принци зниження розмірності

Згідно з прикладом, PCA знаходить спосіб трансформувати двовимірне подання даних в одномірне. Так, замість двох вхідних параметрів x і y, він створює новий параметр k, який є проєкцією з 2d в 1d.

Так, при трансформації відбувається певна втрата даних. На лівому графіку видно, що точки не лежать точно на осі k, але правому спроєктовані точки розміщуються безпосередньо у ній.

Насправді при тисячах вхідних параметрів PCA може скоротити їх кількість у 5-10 разів. Зниження розмірності частіше застосовуються як допоміжний інструмент для алгоритмів машинного навчання з учителем.

2.1.4 Навчання з підкріпленням

Навчання з підкріпленням (Reinforcement Learning, або RL) більшу частину часу працює з цілями безпосередньо штучного інтелекту — створенням агента, який зможе робити ефективні дії в заданому середовищі. Алгоритми RL використовують винагороду як зворотний для виконаних дій і намагаються його максимізувати.

Таке навчання є окремим випадком контрольованого навчання, але вчителем в цьому випадку є «середовище» [38]. Машина не має попередньої інформації про середовище, але має можливість здійснювати в ній будь-які дії. Середа реагує на ці дії та тим самим надає агенту дані, які дозволяють йому реагувати на них і вчитися. Фактично агент і середовище утворюють систему зі зворотним зв'язком (рис 2.8).

Навчання з підкріпленням використовується для вирішення більш складних завдань, ніж навчання з учителем і без вчителя. Воно використовується, наприклад, в системах навігації для роботів, які навчаються уникати зіткнень з перешкодами шляхом набуття досвіду, отримуючи зворотний зв'язок при кожному зіткненні. Основна складність у застосуванні RL у роботі полягає в тому, що реальний світ дуже складно змоделювати з необхідною точністю. Внаслідок цього отриманий штучний інтелект може ідеально виконувати завдання у віртуальному середовищі, але бути практично непридатним у продакшн-умовах.

Рисунок 2.8 — Принцип роботи алгоритму навчання з підкріпленням

На цей час основні дослідження у навчанні із підкріпленням спрямовані на побудову штучного інтелекту для різних класичних відеоігор без опису правил гри [39]. Інакше кажучи, спочатку ШІ нічого не знає про ігрове середовище та знає лише кілька дій. Застосовуючи ці дії, він отримує відгук від гри та модифікує себе через механізм винагород/покарань.

Навчання з підкріпленням використовується також в логістиці, при складанні графіків і плануванні завдань, при навчанні машини логічним іграм.

2.2 Способи використання машинного навчання в кібербезпеці

У великих компаніях аналітики безпеки займаються виявленням шкідливих атак, аналізом мережі, захистом кінцевих точок, оцінкою вразливостей та багатьом іншим. Для цього експертам доводиться оперувати великою кількістю даних, у яких легко проґавити рідкісні, але важливі події. Тут на допомогу приходить машинне навчання. Сфера використання ML у сфері кібербезпеки величезна, починаючи з виявлення аномалій і підозрілої чи незвичайної поведінки та закінчуючи виявленням уразливостей нульового дня та виправленням відомих.

Поєднання різних методів ML підвищує ефективність розпізнавання шкідливого ПЗ та попередження атак. Таким чином, реалізується поведінкова аналітика, наприклад, коли логгується, а потім аналізується послідовність подій у період виконання процесу. Класифікувавши подію, ML-модель зводить його до набору бінарних векторів та навчає глибоку нейронну мережу відрізняти небезпечну активність від логів легітимних подій [40].

Навчання з учителем на сьогодні є найбільш розвиненим і застосовним різновидом машинного навчання. Щоб реалізувати її на практиці, вам потрібне завдання, яке може бути сформульоване у вигляді проблеми класифікаційного або регресійного аналізу, а також достатній набір маркованих даних. Зараз існують десятки готових класичних алгоритмів машинного, а також різні алгоритми глибинного навчання (Deep Learning) для вирішення складніших завдань, таких як обробка зображення, тексту та голосу.

З іншого боку, машинне навчання без вчителя набагато менш застосовне насправді. У той час як асоціативні алгоритми допомагають в аналізі даних для роздрібних та онлайн-магазинів, кластеризація та зниження розмірності більше застосовуються як допоміжний інструмент для алгоритмів машинного навчання з учителем.

Зараз проводиться безліч досліджень із застосування нейронних мереж для розпізнавання складних патернів у немаркованих даних. Потенційно можуть призвести до прориву. Маючи у своєму розпорядженні лише деякими довільними

даними, алгоритми навчання без вчителя можуть бути здатні виявляти деякі нетривіальні залежності або навіть певною мірою складні закони.

Навчання з підкріпленням — дуже перспективний напрямок для вирішення проблем, з якими може впоратися тільки людина. Наразі основні дослідження сконцентровані навколо навчання штучного інтелекту різним видам ігор. Основна перешкода застосування RL на практиці — висока складність реального світу.

Кількість досліджень використання ML алгоритмів в кібербезпеці свідчить про те, що перспектив у цього напрямку багато. Розглянемо приклади застосування технологій ML в кібербезпеці:

- Нейронні мережі застосовуються для виявлення та запобігання вторгненням, але є також пропозиції щодо використання нейронних мереж у «Виявленні відмови в обслуговуванні (DoS), виявленні комп'ютерних хробаків, виявлення спаму, виявлення зомбі, класифікації шкідливих програм та судово-медичних розслідувань» [41]. Такі методи штучного інтелекту, як евристика, інтелектуальний аналіз даних, нейронні мережі та AIS, також були застосовані до антивірусних технологій нового покоління [42].
- Навчання з вчителем дозволить класифікувати нові дані, виявляючи в них щось аномальне, щоб виявити завантаження раніше невідомого шкідливого коду, спам та фішингові атаки, DGA-домени (автоматично створювані шкідливі домени), комунікації з командними серверами та ботнетами.
- Алгоритми класифікації (дерева рішень, випадковий ліс, метод опорних векторів) допоможуть передбачити категорію загрози/уразливості. Таким чином, можна детектувати, наприклад, атаки SQL-Injection або підозрілий трафік.
- Захист від DDoS-атак, особливо від TCP-SYN flood атак та ICMP flood атак здійснюється за допомогою використанням машинного навчання в мережах ISP на основі програмно конфігурованої мережі. Алгоритм машинного навчання, заснований на методі k-найближчих сусідів, що спрощує операції в реальному часі, використовується для виявлення та усунення шкідливого трафіку, відстежуючи джерела атаки, у той час, як нормальний трафік практично не торкається.

- Аналітика поведінки користувачів і об’єктів (UEBA) використовує можливості ML для аналізу журналів поведінки та мережевого трафіку в режимі реального часу та відповідного реагування в разі атаки [43]. Цей процес здійснюється шляхом повторного входу користувача, блокування атаки або оцінки рівня ризику та сповіщення співробітників із інформаційної безпеки компанії, щоб вони могли вжити необхідних заходів.

- Більшість методів ML та DL, таких як навчання ансамблю, кластеризація та дерево рішень, [44]. використовуються для виявлення неправомірного використання, аномалій та гібридного кібервторгнення. За допомогою алгоритмів кластеризації можна виявити витоку інформації внаслідок неправомірних дій користувачів, аналізуючи логи їх поведінки та стан даних.

- Кібератаки, які здійснюють хактивісти, пов’язані з поширеною думкою про резонансні новини. Інформація, зібрана з соціальних мереж, може допомогти передбачити такі інциденти з використанням методів НЛП та ML [45].

- Використання ML для ідентифікації автора програми за допомогою системи, яка може «деанонімізувати» програмістів. [46]. Визначити розробника шкідливого ПЗ тепер набагато простіше, лише аналізуючи вихідний код або скомпільовані двійкові файли [47].

- Методи машинного навчання широко використовуються в системах виявлення вторгнень (IDS) та запобігання вторгненням (IPS) на основі аномалій шляхом навчання моделі для визначення нормальних та аномальних дій. Проблема виявлення вторгнень є класифікацією. Таким чином, алгоритми навчання з учителем часто використовуються для виявлення вторгнень [48].

- Серед методів ML спеціальні прогнозні також можуть бути використані для запобігання втраті/витоку даних (DLP), щоб зменшити ризик злому або витоку [49]. Програмні рішення DLP дозволяють нам встановлювати бізнес-правила, які класифікують конфіденційну та конфіденційну інформацію таким чином, щоб її не можна було розкрити зловмисно або випадково неавторизованими кінцевими користувачами. Цей процес можна здійснити за допомогою алгоритмів навчання під керівництвом і двох типів прикладів: позитивних прикладів (тобто вмісту, який

потрібно захистити) і контрприкладів (тобто документів, які подібні до позитивного набору, але не повинні бути захищені).

Висновок до другого розділу

Машинне навчання стало критичною технологією в інформаційній безпеці, оскільки воно здатне швидко аналізувати мільйони подій і визначати багато типів загроз – від шкідливого програмного забезпечення, що використовує вразливості нульового дня, до виявлення ризикованої поведінки, яка може призвести до фішингової атаки або завантаження шкідливого коду.

Ці технології вчаться з часом, спираючись на минуле, щоб визначити нові типи атак зараз. Історія поведінки створює профілі користувачів, активів і мереж, що дозволяє ML виявляти відхилення від встановлених норм і реагувати на них.

У цьому розділі було розглянуто три напрямки машинного навчання: з учителем, без учителя та з підкріпленням. У кожного є сфери практичного застосування в реальних умовах і власні відмінні риси.

РОЗДІЛ 3
ЗАХИСТ ІНФОРМАЦІЙНИХ РЕСУРСІВ КОРПОРАТИВНОЇ МЕРЕЖІ

3.1 Атаки на алгоритми машинного навчання

Безпека є важливою частиною будь-якої моделі машинного навчання (ML), особливо коли йдеться про ризики, пов'язані з ШІ. Ризики машинного навчання є найголовнішими проблемами, оскільки моделі ML працюють з конфіденційною інформацією, яку необхідно захищати. Ось чому під час побудови моделі ML запобігти атакам і мінімізувати ризик є критичними.

Середовище, в якому виконуються алгоритми машинного навчання, схильна до більшості стандартних векторів атак [50]. Алгоритми машинного навчання схильні до особливих загроз, від яких погано допомагають стандартні заходи захисту.

Зростання поширення машинного навчання та ШІ загалом, ймовірно, буде корелювати зі зростанням ворожих атак. На щастя, існують ефективні заходи безпеки моделі ML, які ви можете включити в рішення DataOps і MLOps, які можуть зупинити атаки до того, як вони відбудуться. Нижче наведено огляд найпоширеніших атак безпеки моделі ML та рішень, які можуть їм запобігти.

3.1.1 Види атак на ML-алгоритми

Атак на ML-алгоритми можна згрупувати за трьома основними векторами.

1. Атаки на дані

Атаки націлені на етапи навчання та перенавчання моделі. Ідеться про псування датасета з метою вплинути на роботу моделі машинного навчання.

Data Poisoning (отруєння даних) передбачає додавання невірних даних, зміну розмітки або перевпорядкування наявних даних. Зрозуміло, недоліки датасета успадковуються моделлю і впливають на її точність і продуктивність [51].

Поверхня подібних атак охоплює весь ланцюжок постачання даних, починаючи з інформаційних брокерів та відкритих датасетів, закінчуючи інформацією, яку система збирає у процесі роботи [52].

Як правило, атаки Data Poisoning спостерігаються у ворожих умовах. Наприклад, лиходії намагаються впливати на роботу спам-фільтрів, щоб поступово знизити їхню ефективність. Проблема ускладнюється тим, що на роботу моделі може вплинути навіть невеликий обсяг надісланих даних. В одному з досліджень це продемонстровано на алгоритмах для прогнозування відсоткових ставок за кредитами, цінами на нерухомість та дозуваннями ліків. Додавши 8% шкідливих навчальних даних, дослідники досягли 75% зміни дозувань, запропонованих алгоритмом для половини пацієнтів [53].

2. Атаки на конвеєр

Model poisoning (отруєння моделі) — втручання в алгоритми, що використовуються у процесі навчання.

Зловмисник може внести зміни до архітектури класифікатора, щоб порушити роботу або маніпулювати ним, а може підмінити модель цілком.

Отруєння моделі може бути наслідком класичної кібератаки, але існує й інший шлях використання слабких місць в алгоритмах машинного навчання.

Цю загрозу демонструють з прикладу федеративного навчання. Коли архітектура моделі включає кілька локальних агентів, компрометація одного з них дозволяє компрометувати модель повністю, погіршити її роботу і впровадити бекдори.

Модель з бекдором працює в штатному режимі, доки не отримає на вхід дані з тригером, який змінює її поведінку так, як потрібно зловмиснику. Наприклад, хакер може обдурити нейронну мережу класифікації зображень, щоб зробити неправильний прогноз, змінивши лише один піксель у вхідному зображенні. Це може призвести до того, що нейронна мережа помилково ідентифікує зображення з певною мірою впевненості.

3. Атаки на розгорнуту модель

Під evasion attack (атака ухилення) мається на увазі зміна вхідних таким чином, що ML модель не може їх правильно ідентифікувати. Як правило, зміни непомітні, але впливають на результати роботи. Такі дані називають “суперечний приклад” adversarial examples (AE).

Найбільш відомий приклад такої атаки пов'язаний із дорожніми знаками [54]. Дослідникам вдалося порушити процес їхнього розпізнавання за допомогою декількох наклейок. Насправді з adversarial examples найчастіше стикаються системи фільтрації контенту, наприклад, алгоритми модерації соціальних мереж.

Складність створення AE, а отже, і доступність подібних атак залежить від того, як багато відомо про алгоритм машинного навчання, який необхідно обдурити. Якщо зловмисник має доступ до моделі, він може порівняно легко створювати різноманітні examples за допомогою методів оптимізації. Крім того, для підготовки до атак можна використовувати датасет, на якому навчалася модель. Для цього необхідно навчити "модель-копію". Створений для неї AE можливо обдурити [55].

При цьому варто мати на увазі, що теоретично і розгорнуту модель, і її навчальну вибірку можна вкрасти [56].

При атаках Model stealing та Data extraction зловмисник аналізує вхідну, вихідну та іншу зовнішню інформацію про систему, щоб відновити за ними параметри або навчальні дані моделі. Дослідники з Інституту інформатики Макса Планка ще у 2018 році показали [57], як отримати інформацію з ML-алгоритму за допомогою послідовності запитів введення-виведення.

Згідно з відкритими джерелами, Model stealing і Data extraction поки що реалізовані тільки в лабораторних умовах, але навіть теоретична можливість подібних атак заслуговує на увагу і вивчення. Вони є особливо небезпечними для алгоритмів, навчених на конфіденційній інформації: медичних записах, фінансових даних.

3.1.2 Протидія атакам

Іноді цілісність датасетів забезпечують за допомогою криптографії, наприклад, хешування та зберігання підписаних електронним цифровим підписом хеш-сум. Щоб оптимізувати цей процес, дослідники рекомендують застосовувати хеш-дерева. Втім, з урахуванням площі атаки, кращим є інший підхід.

У посібнику Microsoft [58] рекомендується вибудовувати систему безпеки машинного навчання, виходячи з того, що дані вже скомпрометовані. У такому разі необхідно виявити потенційно отруєні зразки та виключити їх із навчальної вибірки.

- Один з перших методів вирішення цього завдання - RONI (Reject on negative impact) передбачає чи не простий перебір — виняток із моделі по одній точці даних (data point), перенавчання та порівняння результатів. І так щоразу, що досить повільно і непрактично.
- Data Provenance Based Approach використовує ймовірність отруєння і той факт, що вона вища для точок з одним і тим самим походженням. У цьому методі датасет спочатку кластеризують, спираючись на метадані, а вже потім перенавчають модель. Проте, щоб це спрацювало, інформація про походження точок має бути правдивою.
- Keyed Non-parametric Hypothesis Tests покладається порівняння нових даних з еталонним датасетом [59], який відбиває передбачуване «нормальне» розподіл даних.

Ці підходи порівняно прості в реалізації, але варто пам'ятати, що вони схильні до помилок і можуть пропускати розтягнуті в часі атаки [60].

Виявлення та протидія впровадженню бекдорів (дефект алгоритму, який дозволяє отримати несанкціонований доступ до даних або дистанційного керування операційною системою або комп'ютером)

Data Poisoning це ще й один із методів впровадження бекдорів у модель машинного навчання. Тому вищезазначені заходи знижують ймовірність появи «закладок» в алгоритмі, але з тим існують і спеціалізовані підходи виявлення бэкдоров.

Розробляються і методи мінімізації збитків. Наприклад, досліджується можливість перенавчання моделі за допомогою невеликої підмножини чистих даних [61].

- Протидія атакам ухилення

Засоби захисту від подібних атак орієнтовані на очищення adversarial examples та попереднє навчання моделі.

Щоб звузити поверхню атаки, застосовують перетворення даних. Найбільше таких методів запропоновано для зображень [62]: метод головних компонентів, граничне перетворення, вейвлет-перетворення, зменшення глибини кольору, обрізка та масштабування, але, схоже, що стиснення JPEG є ефективним у найбільшій кількості випадків.

Інша відома техніка – застосування змагального навчання. Вона полягає у навчанні моделі на наборі, доповненому правильно позначеними adversarial examples. Їх можна згенерувати за допомогою атак FGSM та PGD. Використання Projected Gradient Descent дозволяє створювати моделі з досить гарною стійкістю до атак ухилення [63].

Такі методи підвищують надійність моделі з допомогою збільшення витрат навчання. Проте всі вони, включаючи регулярність, сертифіковане навчання, маскування градієнта, мережу дистиляції втрачатимуть ефективність у міру ускладнення атак.

- Протидія крадіжці моделі та даних

Основний засіб протидії подібним атакам - обмеження кількості запитів до моделі, які можна зробити за одиницю часу. Крім того, існують свідчення [64], що правильно складені ансамблі нейромереж ускладнюють реверсинженіринг та дозволяють захистити навчальну вибірку.

Також варто відзначити оригінальний підхід [65], який дозволяє заявити про крадіжку інтелектуальної власності та довести авторські права на модель у суді. Насправді він передбачає впровадження в модель свого бекдора, який змушує її видавати водяні знаки.

- Рішення для DDoS-атак

Для боротьби з цією атакою використовується аналогічний підхід до традиційної DDoS-атаки на вебсервіси. Архітектура обслуговування стійкої виробничої моделі може протистояти раптовим стрибкам запитів на обслуговування. Унікальність програм машинного навчання полягає в тому, що екземпляри висновку часто оптимізовані для машинного навчання. Це може бути дорожчим у масштабуванні, тому слід вживати більше заходів для пом’якшення.

- Трансферне навчання

Щоб захистити від атак навчання передачі, моделі навчання з перенесенням слід навмисно перенавчати для користувацьких наборів даних. Архітектура моделі повинна бути налаштована, а функції об’єктів оновлені, щоб створити достатньо варіацій у виробничій моделі, щоб вона не була настільки вразливою для атак, призначених для цих кореневих моделей з відкритим кодом.

3.1.3 Комплексний захист ML-алгоритмів

ML-програми вимагають розробки нових стратегій захисту, що включають і специфічні технічні заходи та організаційні - впровадження безпечного циклу розробки та експлуатації.

Це особливо важливо для систем, які навчаються на вхідних даних. Така ML-модель, що розвивається, не може бути повністю протестована заздалегідь.

Комплексна безпека ML-алгоритмів складається з контролю вхідних даних, розуміння внутрішньої роботи моделі, постійного моніторингу та аудиту під час роботи.

Зрештою, забезпечення того, щоб дані, які вводяться у модель ML, не були шкідливими за обсягом, частотою та розподілом, є першорядним для отримання правильного результату від моделі ML.

Хоча створення нових моделей машинного навчання на основі чинної архітектури має багато переваг (вартість, зусилля, точність), без використання належних заходів безпеки, це створює вразливості до атак машинного навчання в нових моделях. Моделі ML, здатні до онлайнового або безперервного навчання,

також відкриті для упередженості та дезінформації, і їх потрібно буде ретельно контролювати.

Розуміння та впровадження заходів безпеки моделі ML та безпеки ШІ дають змогу масштабувати та вирішувати ваші найбільші проблеми, мінімізуючи вразливість та ризик.

3.2 Шляхи запобігання витоку інформації за допомогою ML

Отримання негативних результатів використання моделі ML для захисту інформації може свідчити про витік даних. Це одна з основних помилок машинного навчання. Витік даних у машинному навчанні відбувається, коли дані, які ми звикли для навчання алгоритму машинного навчання, містять інформацію, яку модель намагається передбачити, що призводить до ненадійних і поганих результатів прогнозування після розгортання моделі.

3.2.1 Витік даних в ML

У машинному навчанні під витоком даних розуміється помилка, допущена розробником моделі машинного навчання, коли вони випадково обмінюються інформацією між тестовими та навчальними наборами даних.

Як правило, під час поділу набору даних на набори для тестування та навчання мета полягає в тому, щоб ці два набори не поділилися даними. Це пов’язано з тим, що метою тестового набору є моделювання реальних даних, невидимих для певної моделі. Однак, оцінюючи модель, користувач отримує повний доступ як до навчальних, так і до тестових наборів. Виходячи з цього, головний обов’язок — переконатися, що навчальні та тестові дані не перетинаються.

Через витік даних є вірогідність отримати нереально високий рівень продуктивності моделі ML на тестовому наборі, оскільки ця модель запускається на даних, які вже бачила у певній якості в навчальному наборі. Модель ефективно запам'ятовує дані навчального набору і легко може правильно виводити мітки або

значення для цих прикладів тестового набору даних. Очевидно, що це неправильно, оскільки вводить в оману людину, яка оцінює модель. Коли така модель потім використовується для справді невидимих даних, які надходять здебільшого на виробничій стороні, то продуктивність цієї моделі буде значно нижчою, ніж очікувалося, після розгортання.

Простіше кажучи, витік даних відбувається, коли дані, які використовуються в процесі навчання, містять інформацію про те, що модель намагається передбачити. Це схоже на «шахрайство» в системі. Таким чином, витік даних є серйозною і широко поширеною проблемою в аналізі даних і машинному навчанні.

Проблема витоку даних полягає в тому, що при розмежуванні даних на навчальні та тестові підмножини, деякі ваші дані, присутні в наборі тестів, також копіюються в навчальний набір і навпаки.

В результаті, коли ви тренуєте свою модель за допомогою такого типу розбиття, вона дасть дійсно хороші результати на навчальному та тестовому наборі, тобто точність як навчання, так і тестування повинна бути високою. Але коли ви розгортаєте свою модель у виробництві, вона не працюватиме добре, тому що, коли надходить новий тип даних, вона не зможе з цим впоратися.

3.2.2 Приклади витоку даних в ML

Найочевиднішою і легкою для розуміння причиною витоку даних є включення цільової змінної як функції.

Цільова змінна — результат, який машинне навчання намагається передбачити.

Особливості (Features) — дані, які використовуються моделлю для прогнозування цільової змінної.

Нижче наведено приклади ситуацій витоку даних в ML:

- Після включення цільової змінної як функції відбувається те, що мета передбачення була знищена. Ймовірно, це буде зроблено помилково, але під час

моделювання будь-якої моделі ML потрібно переконатися, що цільова змінна відрізняється від набору функцій.

- Щоб правильно оцінити конкретну модель машинного навчання, потрібно поділити доступні дані на навчальні та тестові підмножини. Незмінно трапляється так, що деяка інформація з тестового набору передається навчальному набору, і навпаки. Отже, ще однією поширеною причиною витоку даних є включення тестових даних до даних навчання. Тому виникає необхідність тестувати моделі з новими та раніше небаченими даними.

Виконуючи дослідницький аналіз даних, ми можемо виявити особливості, які дуже сильно корелюють із цільовою змінною. Звичайно, деякі функції корелюють більше, ніж інші, але на диво високу кореляцію потрібно перевіряти та ретельно обробляти. Таким чином, за допомогою аналізу ми можемо досліджувати вихідні дані за допомогою інструментів статистики та візуалізації. Також, якщо після завершення навчання моделі, особливості мають дуже велику вагу, то слід звернути на це увагу. Ці функції можуть бути джерелами витоків даних.

3.3 Розробка рекомендацій щодо запобігання витоку інформації

Розглянуті вище причини витоку даних свідчать про те, що головною причиною є як і коли ми поділяємо наш набір даних. Через це автором цієї роботи були розроблені рекомендації для запобігання витоку інформації

1. Вилучення відповідного набору функцій для моделі.

Вибираючи ознаки, ми повинні переконатися, що дані ознаки не співвідносяться з заданою цільовою змінною, а також що вони не містять інформації про цільову змінну, яка природно недоступна на момент прогнозування (рис 3.1).

Рисунок 3.1 — Вибір найкращого набору функцій для моделі ML

2. Створення окремого набору перевірки

Щоб мінімізувати або уникнути проблеми витоку даних, ми повинні спробувати відкласти набір перевірки на додаток до наборів навчання та тестування (рис. 3.2). Мета набору перевірки — імітувати реальний сценарій, і використання його як останнього кроку. Таким чином, ми визначимо, чи є будь-який можливий випадок переобладнання, що, своєю чергою, може діяти як застереження щодо розгортання моделей, які, як очікується, будуть неефективними у виробничому середовищі.

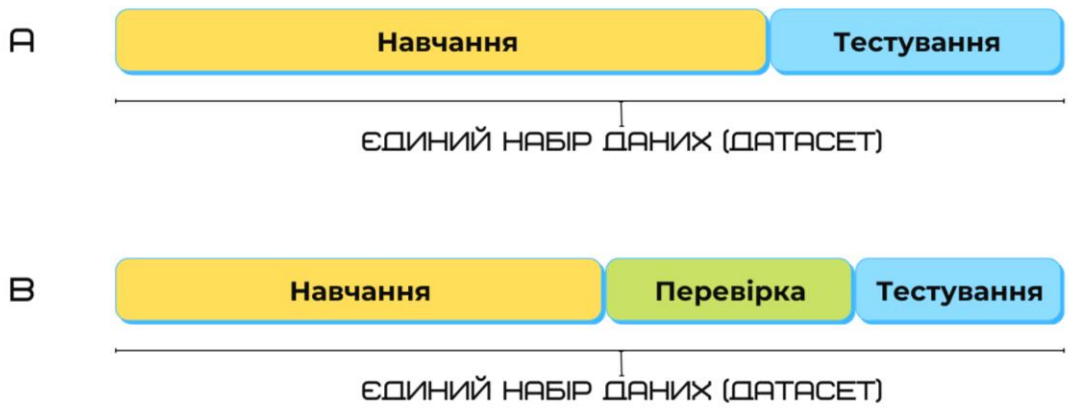

Рисунок 3.2 — Поділ набору даних на підмножини навчання, перевірки та тестування

3. Застосування попередньої обробки даних окремо до підмножин для навчання та тестування

Маючи справу з нейронними мережами, зазвичай нормалізуємо вхідні дані, перш ніж вводити їх у модель. Як правило, нормалізація даних здійснюється шляхом ділення даних на їх середнє значення. Найчастіше нормалізація застосовується до загального набору даних, що впливає на навчальний набір з інформації тестового набору і в кінцевому підсумку призводить до витоку даних. Отже, щоб уникнути витоку даних, потрібно застосувати будь-яку методику нормалізації окремо як до навчальних, так і до тестових підмножин. На рисунку 3.3 зображені розділені дані на навчальні та тестові підмножини.

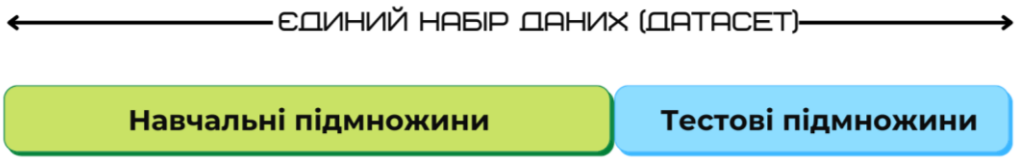

Рисунок 3.3 — Розділені дані на навчальні та тестові підмножини

4. Використання дані тимчасових послідовностей

Маючи справу з типом часових даних, потрібно приділяти більше уваги витоку даних. Таким чином при використанні даних майбутнього, роблячи обчислення для поточних функцій або прогнозів, існує висока ймовірність того, що утвориться витік інформації. Зазвичай це відбувається, коли дані випадковим чином розбиваються на навчальні та тестові підмножини.

Тому під час роботи з даними часових рядів ми встановлюємо граничне значення за часом, що може бути дуже корисним, оскільки не дає нам отримати будь-яку інформацію після часу передбачення (рис 3.4).

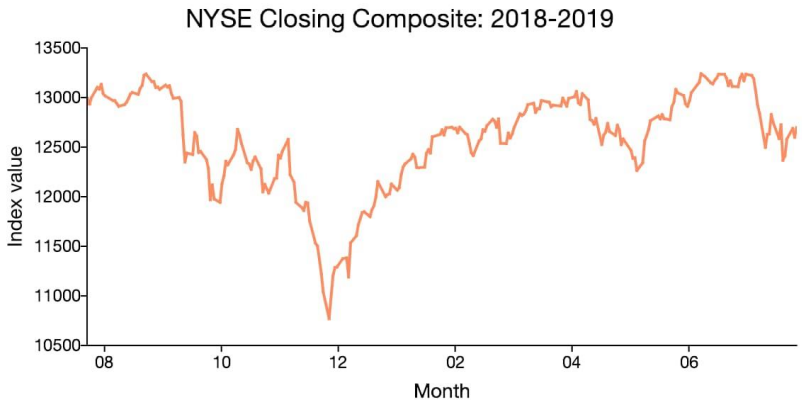

Рисунок 3.4 — Приклад даних тимчасових послідовностей

5. Перехресна перевірка

Якщо існує обмежена кількість даних для навчання алгоритму машинного навчання, доцільно використовувати перехресну перевірку в процесі навчання. Перехресна перевірка полягає в тому, що вона розбиває наші повні дані на k складок і повторює весь набір даних k кількість разів, і кожен раз використовуємо k-1 раз для навчання та 1 раз для тестування нашої моделі (рис 3.5).

Рисунок 3.5 — Концепція перехресної перевірки

Перевага цього підходу полягає в тому, що під час перевірки ми використовуємо увесь набір даних як для навчання, так і для тестування. Однак для запобігання витоку даних краще масштабувати або нормалізувати дані та обчислювати параметри для кожної частини перехресної перевірки окремо.

Висновок до третього розділу

Витік даних є широко поширеною проблемою в корпоративних мережах. Щоб модель машинного навчання мала хороші показники в цих прогнозах, потрібно приділяти пильну увагу виявлянню та уникненню витоку даних.

Машинне навчання — це розробка шаблонів і маніпулювання цими шаблонами за допомогою алгоритмів. Щоб розробити шаблони, потрібно багато даних звідусіль, тому що дані повинні представляти якомога більше потенційних результатів із якомога більшої кількості потенційних сценаріїв.

Машинне навчання має безліч переваг та недоліків. Деякі алгоритми уразливі, особливо при розподілу даних. Через це були розроблені рекомендації для запобігання витоку інформації.

ВИСНОВКИ

У дипломній роботі розв’язано актуальне наукове завдання щодо порівняння систем для запобігання витоку інформації в корпоративних мережах.

Завдяки машинному навчанню системи кібербезпеки можуть аналізувати закономірності та вчитися на них, щоб запобігти подібним атакам і реагувати на зміну поведінки. Це допоможе фахівцям з кібербезпеки бути більш активними у запобіганні загроз і реагуванні на активні атаки в режимі реального часу. Методи ML оптимізують кількість часу, витраченого на рутинні завдання, через це організації використовують власні ресурси більш стратегічно.

Машинне навчання може зробити кібербезпеку простішою, активнішою, менш витратною та набагато ефективнішою. Але повній відмові від колишніх методів кібербезпеки на користь машинного навчання перешкоджають вже зазначені причини:

- відсутність достатньої кількості датасетів для коректного навчання ML-моделей у сфері кіберзагроз;
- можливість специфічних атак на ML-алгоритми та використовувані датасети, що може призвести до неправильних рішень, пропущених атак або помилкових спрацьовувань;
- зловмисники теж використовують алгоритми ML для створення шкідливих програм, аналізу користувача поведінки, розробки ботів-збирачів персональних даних, пошуку вразливостей, підбору паролів, підміни особистості, обходу систем захисту тощо.

Проведений аналіз систем захисту корпоративних мереж IDS, IPS, DLP, SIEM, NBAD дозволив зробити висновок, що застосування машинного навчання може значно доповнити та розширити функціонал наявних систем. Зокрема, ML-моделі підвищують точність сигнатурного аналізу, який швидко обробляє запити та не вимагає тривалого періоду навчання. Таким чином, можна використовувати сигнатурний аналіз для виявлення запитів з явними ознаками атак, а машинне

навчання – для аналізу інших запитів. У результаті такого поєднання різних методів досягається висока швидкість роботи антивірусного ПЗ з мінімальною кількістю пропусків атак.

У роботі були детально проаналізовані види атак на алгоритми машинного навчання та протидія ним, досліджено шляхи запобігання витоку інформації за допомогою ML та розроблено п'ять рекомендацій для запобігання витоку даних. Дані рекомендації дозволять вирішити проблему витоку даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Cyber Security Statistics The Ultimate List Of Stats, Data & Trends [Електронний ресурс], Purplesec. – 2022. – Режим доступу до ресурсу: https://purplesec.us/resources/cyber-security-statistics/.

2. John P. M. Data Breaches Affected Nearly 6 Billion Accounts in 2021 [Електронний ресурс], Mello Jr John P. Tech News World. – 2022. – Режим доступу до ресурсу: https://www.technewsworld.com/story/data-breaches-affected-nearly-6-billion-accounts-in-2021-87392.html.

3. Sharma, B. Mangrulkar, R. (2019). Deep learning applications in cybersecurity: a comprehensive review, challenges and prospects. International Journal of Engineering Applied Sciences and Technology, 4(8), 148-159

4. Chapple M. Stewart J. M. Gibson D. (ISC) 2 CISSP Certified Information Systems Security Professional Official Study Guide. – John Wiley & Sons, 2018. 12.

5. Предотвращение утечек данных – DLP [Електронний ресурс]. – Режим доступу: http://allta.com.ua/nashi-resheniya/informacionnayabezopasnost/dlp-systems.

6. Загальні положення щодо захисту інформації в комп’ютерних системах від несанкціонованого доступу: НД ТЗІ 1.1–002–99.–К.. ДСТСЗІ СБ України, 1999. – 16 с.

7. Типи загроз ІБ дуже різноманітні та мають безліч класифікацій — ISO/IEC 27035:2011 «Information technology. Security techniques. Information security incident management».

8. CMU/SEI-2004-TR-015 «Defining incident management processes for CISRT».

9. NIST SP 800-61 «Computer security incident handling guide»

10. Захист від витоків даних – DLP–рішення [Електронний ресурс], Спосіб доступу: URL: http://www.protectme.ru/infosec/dlp.

11. ISO/IEC 27002:2013 «Information technology. Security techniques. Code of practice for information security controls».

12. Сравнительный обзор средств предотвращения утечек данных (DLP) [Електронний ресурс]. – Режим доступу: https://safesurf.com/specialists/article/5233/609990/ – Заголовок з екрана. 10.

13. Внедрение DLP-систем [Електронний ресурс]. – Режим доступу: https://techexpert.ua/our-services/implementation-of-dlp-systems/ – Заголовок з екрана. 11.

14. D. K. Barman, G. Khataniar, (2012) “Design Of Intrusion Detection System Based On Artificial Neural Network And Application Of Rough Set”, International Journal of Computer Science and Communication Networks, Vol. 2, No. 4, pp. 548-552.

15. John P. M. Data Breaches Affected Nearly 6 Billion Accounts in 2021 [Електронний ресурс], Mello Jr John P. Tech News World. – 2022. – Режим доступу до ресурсу: https://www.technewsworld.com/story/data-breaches-affected-nearly-6-billion-accounts-in-2021-87392.html.

16. E. Tyugu, (2021) “Artificial intelligence in cyber defense”, 3rd International Conference on Cyber Conflict (ICCC 2011), pp. 1–11

17. Jibilian I. SolarWinds cyber attack [Електронний ресурс], I. Jibilian, K. Canales Business Insider. – 2020. – Режим доступу до ресурсу: https://www.businessinsider.com/solarwinds-hack-explained-government-agencies-cyber-security-2020-12.

18. Сайт компанії Gartner [Електронний ресурс]. – Режим доступу: https://www.gartner.com/. 15.

19. DLP-системы: защита от утечки информации [Електронний ресурс]. – Режим доступу: http://pro-spo.com/personal-data-security/3738-dlp-sistemy-zashhitaot-utechki-informaczii – Заголовок з екрана. 17.

20. Романюков М.Г. Категоріювання інформації у сучасній структурі кібербезпеки держави з використанням матриць цінностей, М.Г. Романюков. – Харків: Критичні комп’ютерні технології та системи: науково-технічний семінар. 23 травня 2019 року. Тема семінару – Безсерверні архітектури, хмарні технології та кібербезпека. 18.

21. Johansen G. Digital forensics and incident response: an intelligent way to respond to attacks. – 2020.

22. Enterprise Data Loss Prevention (DLP) Reviews and Ratings [Електронний ресурс], Gartner Peer Inside. – 2021. – Режим доступу до ресурсу: https://www.gartner.com/reviews/market/enterprise-data-loss-prevention.

23. Ушаков В. Проблеми оперативного виявлення і реагування на інциденти інформаційної безпеки, В. Ушаков, О. Сєвєрінов, GLOBAL CYBER 69 SECURITY FORUM. Матеріали першого міжнародного науково-практичного форуму – Х.: ХНУРЕ, 2020. – С. 104-105.

24. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. Courville. (2017). Improved training of wasserstein GANs. In Proc. of the 31st International Conference on Neural Information Processing Systems, (pp.5769-5779).

25. Slipachuk L., Toliupa S., Nakonechnyi V. The Process of the Critical Infrastructure Cyber Security Management using the Integrated System of the National Cyber Security Sector Management in Ukraine, 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT). – IEEE, 2019. – С. 451-454.

26. Закон України Про інформацію (Відомості Верховної Ради України (ВВР), 1992, № 48, ст.650)

27. НД ТЗІ 1.1-002-99 Загальні положення щодо захисту інформації в комп'ютерних системах від несанкціонованого доступу

28. Top Security and Risk Management [Електронний ресурс], Gartner. – 2021. – Режим доступу до ресурсу: https://www.gartner.com/en/doc/738210-top-security-and-risk-management-trends-2021.

29. Moore S. 7 Top Trends in Cybersecurity for 2022 [Електронний ресурс] / Susan Moore // Gartner. – 2022. – Режим доступу до ресурсу: https://www.gartner.com/en/articles/7-top-trends-in-cybersecurity-for-2022.

30. E. Tyugu, (2018) “Artificial intelligence in cyber defense”, 3rd International Conference on Cyber Conflict (ICCC 2018), pp. 1–11.

31. B. Stahl, D. Elizondo, M. C. Mayer, Y. Zheng, K. Wakunuma, (2010) “Ethical and Legal Issues of the Use of Computational Intelligence Techniques in Computer Security and Computer Forensics”, International Joint Conference on Neural Networks (IJCNN), pp. 1 8. .

32. https://www.sas.com/en_us/insights/analytics/machine-learning.html#:~:text=Machine%20learning%20is%20a%20method,decisions%20with%20minimal%20human%20intervention.

33. Le Roux, N., Bengio, Y. (2008). Representational power of restricted Boltzmann machines and deep belief networks. Neural computation, 20(6), 1631-1649.

34. Machine Learning [Електронний ресурс], IT Enterprise. – 2021. – Режим доступу до ресурсу: https://www.it.ua/knowledge-base/technology-innovation/machine-learning.

35. Вступ до алгоритмів / Т.Кормен, Ч. Лейзерсон, Р. Рівест, К. Стан. – Київ: К.І.С. 2019. – 1288 с. – (Третє).

36. The Growing Role of Machine Learning in Cybersecurity [Електронний ресурс], Al Perlman Security round table. – 2021. – Режим доступу до ресурсу:https://www.securityroundtable.org/the-growing-role-of-machine-learning-in-cybersecurity/

37. Жураковський Б. ТЕХНОЛОГІЇ ІНТЕРНЕТУ РЕЧЕЙ, Б. Жураковський, І. Зенів. Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського». – 2021. – С. 118.

38. Гайдур Г. ЦИФРОВА ТРАНСФОРМАЦІЯ КІБЕРБЕЗПЕКИ, Г. Гайдур, Г. Найман. Державний університет телекомунікацій. – 2020. – С. 153.

39. Гахов С.О. Гаркавенко Д.М ЦИФРОВА ТРАНСФОРМАЦІЯ КІБЕРБЕЗПЕКИ, Г. Гайдур, Г. Найман. Державний університет телекомунікацій. – 2020. – С. 167.

40. Using Artificial Intelligence in Cybersecurity [Електронний ресурс], Balbix. – 2022. – Режим доступу до ресурсу: https://www.balbix.com/insights/artificial-intelligence-in-cybersecurity/.

41. E. Tyugu, (2011) “Artificial intelligence in cyber defense”, 3rd International Conference on Cyber Conflict (ICCC 2011), pp. 1–11

42. (X. B. Wang, G. Y. Yang, Y. C. Li, D. Liu, (2008) ”Review on the application of Artificial Intelligence in Antivirus Detection System”, IEEE Conference on Cybernetics and Intelligent Systems, pp. 506 509.

43. Chris Brook. What is User and Entity Behavior Analytics? A Definition of UEBA, Benefits, How It Works, and More. Accessed: Oct. 10, 2019. [Online]. Available at: https://digitalguardian.com/blog/what-user-and-entity-behavior-analytics-definition-ueba-benefitshow—it-works-and-more

44. Anna L. Buczak, Erhan Guven. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Communications Surveys & Tutorials, vol. 18, no. 2, 2016, pp. 1153-1176

45. Hernandez-Suarez, G. Sanchez-Perez, K. Toscano-Medina, V. Martinez-Hernandez, H. Perez-Meana, J. Olivares-Mercado, V. Sanchez. Social Sentiment Sensor in Twitter for Predicting Cyber-Attacks Using ℓ1 Regularization. Sensors, vol. 18, no. 5, 2018, pp. 1380.

46. Caliskan, F. Yamaguchi, E. Dauber, R. Harang, K. Rieck, R. Greenstadt, A. Narayanan. Deanonymizing Programmers via Code Stylometry. In Proc. of the 24th USENIX Security Symposium, 2015, pp. 255-270.

47. Caliskan, F. Yamaguchi, E. Dauber, R. Harang, K. Rieck, R. Greenstadt, A. Narayanan. When Coding Style Survives Compilation: De-anonymizing Programmers from Executable Binaries. arXiv:1512.08546, 2015

48. D. Dasgupta, (2006) “Computational Intelligence in Cyber Security”, IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS 2006), pp. 2–3

49. Introduction to Forcepoint DLP Machine Learning. Accessed: Oct. 10, 2019. [Online]. Available at: https://www.websense.com/content/support/library/data/v84/machine_learning/machine_learning.pdf

50. S. Repalle, V. Kolluru. Intrusion Detection System using AI and Machine Learning Algorithm. International Research Journal of Engineering and Technology (IRJET), vol. 04, issue 12, 2017, pp. 1709-1715.

51. Corey D. Robustness Evaluations of Sustainable Machine Learning Models Against Data Poisoning Attacks in the Internet of Things [Електронний ресурс], D. Corey, M. Nour, B. Turnbull, MDPI. – 2020. – Режим доступу до ресурсу: https://www.researchgate.net/publication/343560652_Robustness_Evaluations_of_Sustainable_Machine_Learning_Models_Against_Data_Poisoning_Attacks_in_the_Internet_of_Things.

52. Microsoft deletes racist genocidal tweets from ai chatbot tay [Електронний ресурс], Business insider. – 2016. – Режим доступу до ресурсу: https://www.businessinsider.com/microsoft-deletes-racist-genocidal-tweets-from-ai-chatbot-tay-2016-3?r=UK&IR=T

53. Robust Physical-World Attacks on Deep Learning Models [Електронний ресурс], Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Cornell University. – 2018. – Режим доступу до ресурсу: https://arxiv.org/abs/1707.08945

54. Transferability in Machine Learning: from Phenomena to Black-Box Attacks using Adversarial Samples [Електронний ресурс], Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Cornell University. – 2016. – Режим доступу до ресурсу: https://arxiv.org/abs/1605.07277

55. Stealing Machine Learning Models via Prediction APIs [Електронний ресурс], Florian Tramèr, Fan Zhang, Ari Juels, Cornell University. – 2016. – Режим доступу до ресурсу: https://arxiv.org/abs/1609.02943

56. Towards Reverse-Engineering Black-Box Neural Networks [Електронний ресурс], Seong Joon Oh, Max Augustin, Bernt Schiele, Mario Fritz Cornell University. – 2018. – Режим доступу до ресурсу: https://arxiv.org/abs/1711.01768

57. Protecting the integrity of the training procedure of neural networks [Електронний ресурс], Christian Berghoff, Cornell University. – 2020. – Режим доступу до ресурсу: https://arxiv.org/abs/2005.06928.

58. Failure Modes in Machine Learning [Електронний ресурс], Microsoft. – 2022. – Режим доступу до ресурсу: https://docs.microsoft.com/en-us/security/engineering/failure-modes-in-machine-learning

59. Protecting Machine Learning from Poisoning Attacks [Електронний ресурс], Yao Cheng, Cheng-Kang Chu, Springer. – 2019. – Режим доступу до ресурсу:https://link.springer.com/chapter/10.1007/978-3-030-36938-5_39

60. Robustness Evaluations of Sustainable Machine Learning Models Against Data Poisoning Attacks in the Internet of Things [Електронний ресурс], Corey Dunn, Nour Moustafa, Benjamin Peter Turnbull, Researchgate. – 2020. – Режим доступу до ресурсу: https://www.researchgate.net/publication/343560652_Robustness_Evaluations_of_Sustainable_Machine_Learning_Models_Against_Data_Poisoning_Attacks_in_the_Internet_of_Things

61. Neural Trojans [Електронний ресурс], IEEE. – 2017. – Режим доступу до ресурсу: https://ieeexplore.ieee.org/document/8119189.

62. Defending against Adversarial Images using Basis Functions Transformations [Електронний ресурс], Uri Shaham, James Garritano, Yutaro Yamada, Cornell University. – 2020. – Режим доступу до ресурсу: https://arxiv.org/abs/1803.10840v3

63. Dowards Deep Learning Models Resistant to Adversarial Attacks [Електронний ресурс], Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Cornell University. – 2020. – Режим доступу до ресурсу: https://arxiv.org/abs/1706.06083

64. Stealing Machine Learning Models via Prediction APIs [Електронний ресурс], Florian Tramèr, Fan Zhang, Ari Juels, Cornell University. – 2020. – Режим доступу до ресурсу: https://arxiv.org/abs/1609.02943

65. Jialong Z. Protecting Intellectual Property of Deep Neural Networks with Watermarking / Z. Jialong, J. Jiyong, G. Zhongshu. // IBM Research. – 2021. – С. 9–13.

ДОДАТОК А

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ

Тези наукових доповідей

1. Кравченко Л.Д., Наконечний В.С. Використання машинного навчання для запобігання витоку інформації в корпоративних мережах. V Міжнародна науково-практична конференція “Проблеми кібербезпеки інформаційно-телекомунікаційних систем”, PCSITS 2022, Київ.