https://doi.org/10.31891/2219-9365-2025-83-32
УДК 004.056:004.852

КОНОТОПЕЦЬ Микола
Інститут спеціального зв'язку та захисту інформації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»
https://orcid.org/0000-0002-6963-1877
e-mail: nikolyalux@gmail.com
ТУРОВСЬКИЙ Олександр
Державний університет інформаційно-комунікаційних технологій
https://orcid.org/0000-0002-4961-0876
e-mail:s19641011@ukr.net
БУРДЕЙНИЙ Андрій
Інститут спеціального зв'язку та захисту інформації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»
https://orcid.org/0009-0005-4058-6792
e-mail: burdes228@gmail.com
СТОРЧАК Антон
Інститут спеціального зв'язку та захисту інформації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»
https://orcid.org/0000-0002-5267-3122
e-mail: storchakanton@gmail.com

ПОРІВНЯЛЬНИЙ АНАЛІЗ ЕФЕКТИВНОСТІ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ КІБЕРІНЦИДЕНТІВ

У статті проведено порівняльний аналіз сучасних методів машинного навчання (керованого, некерованого та навчання з підкріпленням) для виявлення кіберінцидентів у корпоративних інформаційно-комунікаційних системах. Розглянуто переваги та обмеження найбільш поширених алгоритмів, зокрема Decision Tree, Naive Bayes, SVM, Isolation Forest, K-Means, BERT, GPT, DQN, PPO та Soft Actor–Critic, у контексті точності, повноти, прецизійності та рівня хибнопозитивних спрацювань. Для експериментальної перевірки використано датасет CICIDS 2018, що дозволило оцінити практичну придатність зазначених методів для виявлення як відомих загроз, так і атак нульового дня. У ході дослідження встановлено, що моделі дерев рішень демонструють найвищу точність і мінімальний рівень хибнопозитивних спрацювань для класичних загроз, тоді як алгоритм Isolation Forest є найбільш ефективним для виявлення аномальної активності та нових типів атак. Запропоновано оптимізований підхід, що поєднує контрольоване навчання (Decision Trees) для виявлення класичних загроз та неконтрольоване виявлення аномалій (Isolation Forest) для мінімізації хибнопозитивних спрацювань і забезпечення адаптивності системи. Отримані результати можуть бути використані для побудови ефективних систем кіберзахисту, здатних оперативно реагувати на сучасні загрози з урахуванням ресурсних обмежень та потреби у зниженні рівня помилкових спрацювань. Окрему увагу приділено оцінці впливу параметрів моделей на їх продуктивність та здатність до масштабування у середовищах з високою інтенсивністю трафіку. Розглянуто можливості інтеграції машинного навчання з існуючими системами моніторингу безпеки для автоматизації виявлення інцидентів. Визначено напрями подальших досліджень, зокрема розробку гібридних моделей для підвищення стійкості до атак нульового дня. Зроблено висновок про доцільність застосування машинного навчання як ключового елементу сучасних стратегій кіберзахисту.

Ключові слова: методи машинного навчання, виявлення кіберінцидентів, забезпечення кібербезпеки, аналіз аномалій, кероване МН, некероване МН, МН з підкріпленням.

KONOTOPETS Mykola, BOURDEINNY Andriy, STORCHAK Anton
Institute of Special Communications and Information Protection National Technical University of Ukraine"Igor Sikorsky Kyiv Polytechnic Institute"
TUROVSKY Oleksandr
State University of Information and Communication Technologies

COMPARATIVE ANALYSIS OF THE EFFECTIVENESS OF MACHINE LEARNING METHODS FOR CYBER INCIDENT DETECTION

The article presents a comparative analysis of modern machine learning methods (supervised, unsupervised, and reinforcement learning) for detecting cybersecurity incidents in corporate information and communication systems. The advantages and limitations of the most common algorithms, including Decision Tree, Naive Bayes, SVM, Isolation Forest, K-Means, BERT, GPT, DQN, PPO, and Soft Actor–Critic, are discussed in terms of accuracy, recall, precision, and false positive rate. The CICIDS 2018 dataset was used for experimental evaluation, allowing the practical applicability of these methods for detecting both known threats and zero-day attacks to be assessed. The study found that decision tree models demonstrate the highest accuracy and the lowest false positive rates for conventional threats, while the Isolation Forest algorithm is the most effective for detecting anomalous activity and new types of attacks. An optimized approach is proposed, combining supervised learning (Decision Trees) for detecting known threats with unsupervised anomaly detection (Isolation Forest) to minimize false positives and enhance system adaptability. The results obtained can be used to build efficient cybersecurity systems capable of promptly responding to modern threats while considering resource constraints and the need to reduce false positives. Particular attention is given to assessing the impact of model parameters on their performance and scalability in high-traffic environments. The possibilities of integrating machine learning with existing security

monitoring systems for incident detection automation are considered. Directions for future research are identified, including the development of hybrid models to increase resilience to zero-day attacks. The study concludes that machine learning should be considered a key component of modern cybersecurity strategies.

Keywords: machine learning methods, cyber incident detection, cybersecurity, anomaly analysis, supervised machine learning, unsupervised machine learning, reinforced machine learning.

Стаття надійшла до редакції / Received 26.07.2025
Прийнята до друку / Accepted 18.08.2025

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Аналітичні звіти [1 – 7] демонструють зростання складних кібератак, несанкціонованих порушень цілісності інформації та викликаних ними кіберінциндентів різної складності. Традиційні методи виявлення кіберінцидентів швидко втрачають актуальність, потребуючи постійного вдосконалення та адаптації до новітніх кібератак. Сигнатурні методи не здатні виявляти атаки нульового дня, а поведінкові методи мають недолік у вигляді великої кількості помилкових спрацювань. В даних обставинах зростає роль та важливість пошуку нових технологій та заснованих на них методах виявлення та аналізу кіберінциндетів. Одним з таких методів є метод виявлення мережевих кібератак на основі застосування нейронних мереж, щозаснований на технології машинного навчання (метод МН).

Метод МН демонструють значний потенціал в сфері виявлення мережевих кіберінциндентів, адже мають здатність самостійно навчатись та адаптуватись до змін в кіберпросторі, що дозволяє досить ефективно виявляти не тільки існуючі але й нові атаки, в тому числі атаки нульового дня.

Результати наукових досліджень щодо застосування методів МН [8-16] показав, що більшість з них описують окремі методи МН, але не надають комплексного порівняння їх ефективності, особливо для виявлення інцидентів в реальному часі та виявлення атак нульового дня.

В Україні питання захисту від атак на інформаційні системи критичної інфраструктури, урядового та енергетичного секторів постає досить болюче. Кібератаки країни–агресора завдають значної шкоди, забираючи життя та дестабілізуючи суспільство, шляхом підриву довіри до уряду та сектору безпеки та оборони. Актуальним залишається дослідження та вибір оптимальних методів МН для забезпечення належного рівня кібербезпеки.

АНАЛІЗ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

Питанню аналізу ефективності методів МН для виявлення кіберінцидентів присвячені роботи різних авторів. Так в роботах [8, 9] описано декілька методів та алгоритмів МН для застосування в сфері кібербезпеки. Також, описано їх особливості, переваги та недоліки в контексті захисту ІКС.

Публікація [10] досліджує застосування МН в кібербезпеці, зокрема для виявлення фішингу, вторгнень в систему та спаму. Автор описує методику, що включає основні етапи: збирання даних, навчання моделей, оцінка та інтеграція в реальні ІКС.

Стаття [11] надає огляд основних алгоритмів МН, їхнього призначення та принципів роботи. Дослідження охоплює усі типи МН, але не описує оптимального рішення для використання, хоча наголошує на автоматизації процесів після навчання та зростаючому попиті на ці технології в медицині, військовій справі та інших галузях.

У [12] описано, як МН аналізує журнали безпеки для автоматичного виявлення атак, підкреслюючи проблему хибнопозитивних спрацювань у великих SOC.

Наукова робота [13] досліджує фактори, що впливають на впровадження алгоритмів МН для виявлення кіберзагроз у банківській сфері. Проаналізувавши матеріал роботи можна дійти висновку, що принципи, описані в цій роботі можуть використовуватись не лише в банківській сфері, а й в багатьох інших, де важливою складовою кібербезпеки є вплив людського фактору.

Публікації [14, 15] показують, як методи МН можуть покращити кібербезпеку в ІКС. Розглянуто одні з основних сучасних методів МН та описано їх вплив у протистоянні сучасним кібератакам. Дослідження підкреслюють ефективність МН на рівнях апаратного забезпечення, мережі та додатків, демонструючи його позитивний вплив на покращення кібербезпеки завдяки розпізнаванню шаблонів і прогнозуванню атак.

Аналіз сучасних досліджень свідчить про активне дослідження методів МН в сфері кібербезпеки. Однак, попри широкий огляд методів, існує низка суттєвих прогалин. Зокрема, недостатньо уваги приділяється аналізу основного недоліку МН – високому рівню помилкових спрацювань, які суттєво знижують ефективність їх використання в умовах реальних ІКС. Крім того, відсутні комплексні дослідження, які б систематизували найпоширеніші методи МН, оцінили їх ефективність з точки зору хибних спрацювань та запропонували оптимальне рішення для практичного впровадження в кіберзахист. Це вказує на необхідність проведення цільових досліджень, орієнтованих на вирішення зазначених питань.

ISSN 2219-9365

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Метою роботи є проведення порівняльного аналізу методів МН для виявлення кіберінцидентів в ІКС. Для досягнення мети роботи необхідно:

1. Оцінити ефективність сучасних методів МН у виявленні кіберінцидентів;
2. Оцінити ймовірність помилкових спрацювань сучасних методів МН;
3. Визначити оптимальний метод МН для подальшого дослідження, вдосконалення та впровадження у корпоративних інформаційних системах з підвищеними вимогами до безпеки.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

На сьогоднішній день МН використовується в кібербезпеці для виявлення аномалій у мережевому трафіку, активності користувачів та системних журналах, а також для виявлення фішингових атак і атак шкідливого програмного забезпечення (ШПЗ), навіть якщо вони не мають відомих сигнатур. Крім того, МН застосовується для оцінки вразливостей, автоматизованої реакції на інциденти та моніторингу активності користувачів з метою виявлення аномальних подій у системі.

Опитування IBM [16] показує, що організації, які не використовували штучний інтелект (ШІ) та МН, витратили в середньому 5,72 млн доларів США на усунення інцидентів, тоді як ті, що широко використовували ШІ та МН мали середні витрати у розмірі 3,84 млн доларів США, зекономило їм 1,88 млн доларів США. Крім того, організації, які широко використовують ШІ та МН в галузі безпеки, виявляли та локалізували порушення безпеки даних майже на 100 днів швидше, ніж організації, які взагалі не використовували ці технології [16].

Все це можливо завдяки аналізу великих наборів даних про зловмисну діяльність та виявленню закономірностей і аномалій.

Машинне навчання є напрямом ШІ, що зосереджується на наданні комп'ютерам і машинам можливості імітувати процес навчання людини, самостійно виконувати завдання та покращувати свою продуктивність і точність завдяки накопиченому "досвіду" та обробці великих обсягів даних [17]. Алгоритми МН дають змогу системам автоматично навчатися на основі даних без явного програмування, що робить їх надзвичайно ефективними в динамічних і складних середовищах [18]. Машинне навчання класифікується за методами, наведеними на рис. 1.

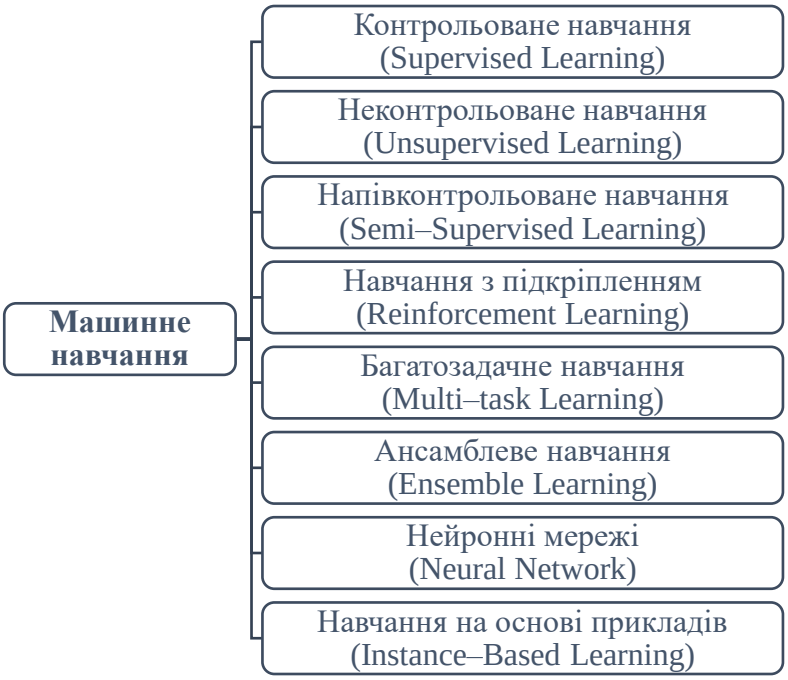

Рис. 1. Методи машинного навчання

З урахуванням поточних тенденцій у сфері кібербезпеки та на основі аналізу актуальних наукових робіт [8-15], можна виокремити три найбільш перспективні методи, наведені на рис. 2. У цій статті розглянуто лише вищевказані методи та їхні алгоритми з позиції їх потенційного використання та розвитку в галузі кіберзахисту, оскільки машинне навчання не стоїть на місці і постійно розвивається у відповідь на появу нових способів атак.

ISSN 2219-9365

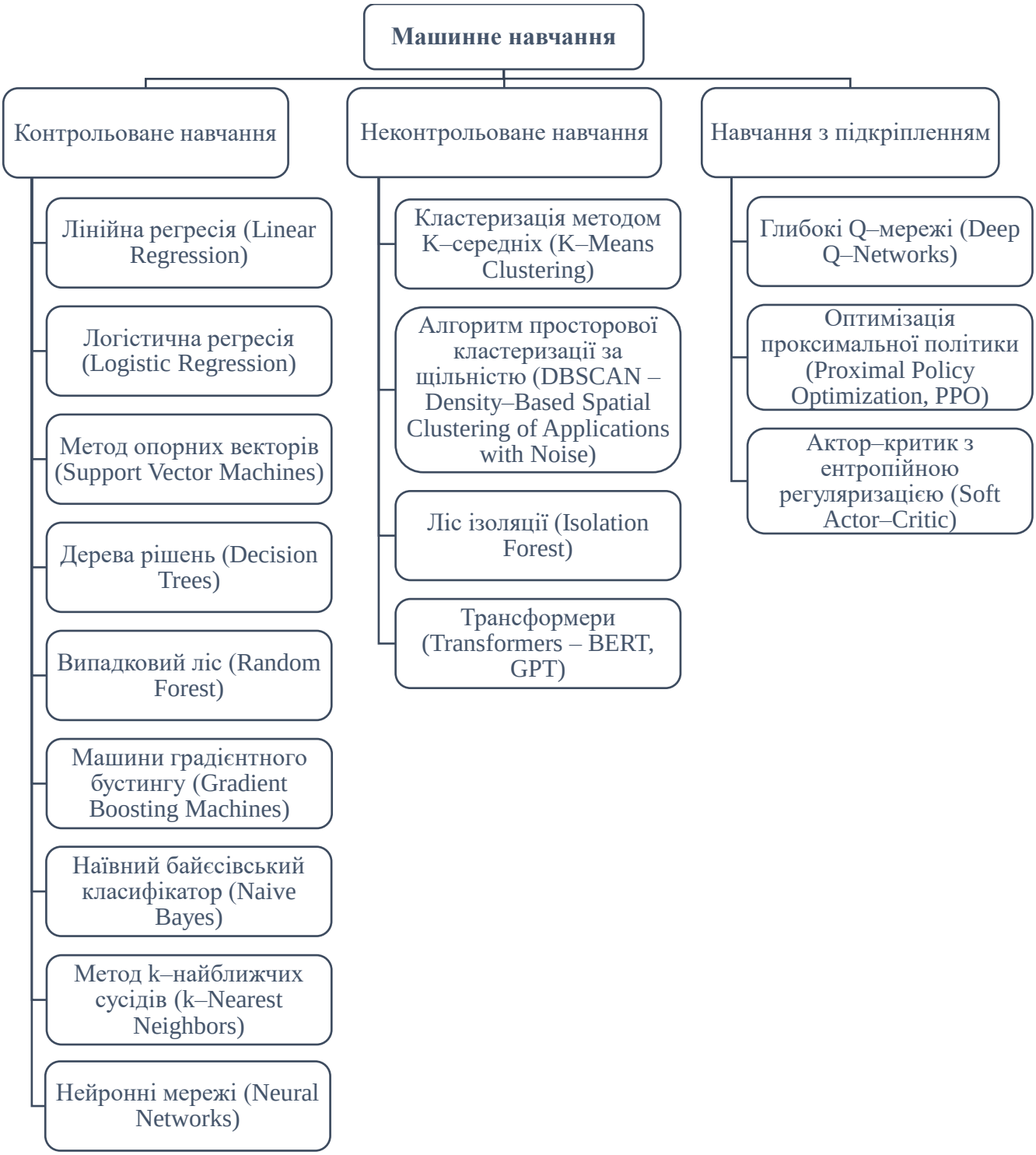

Рис. 2. Перспективні методи машинного навчання

У контрольованому навчанні (Supervised Learning) модель МН навчається на позначеному наборі даних, де надаються пари вхід–вихід, щоб навчити модель робити прогнози. Наразі ця модель домінує в сфері кіберзахисту та є особливо ефективною для виявлення та класифікації ШПЗ [19]. Контрольована модель може бути навчена на наборі даних відомих сигнатур ШПЗ та нешкідливого програмного забезпечення, що дозволяє їй класифікувати нові файли як шкідливі або безпечні на основі вивчених шаблонів.

Кожна точка даних у навчальному наборі має як вхідні характеристики (наприклад, заголовок електронного листа, текст електронного листа, тощо), так і відповідну цільову мітку (наприклад, “спам” або “не спам”). Модель навчається зіставляти вхідні дані з вихідними даними, а потім може прогнозувати мітки для некласифікованих даних.

Контрольоване МН вирішує два типи задач (рис. 3):

- класифікація, яка прогнозує категорії (наприклад, електронний лист є спамом чи ні);
- регресія, яка прогнозує безперервні значення (наприклад, кількісно оцінює рівень ризику, пов'язаний з виявленою вразливістю).

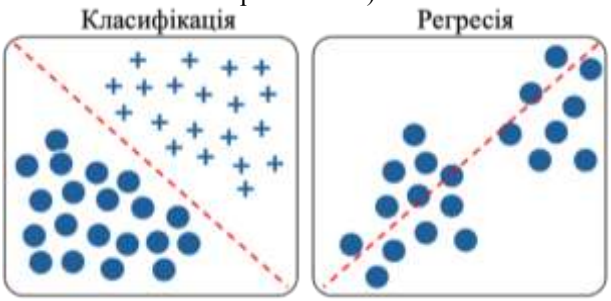

Рис. 3. Задачі МН

Серед усіх алгоритмів контрольованого МН в сфері кібербезпеки найчастіше застосовуються дерева рішень (Decision Trees), наївний байєсівський класифікатор (Naive Bayes) та метод опорних векторів (Support Vector Machines).

Дерева рішень (Decision Trees) – це алгоритм контрольованого МН, що нагадує дерева, які групують атрибути, сортуючи їх за значеннями. Дерева рішень використовують головним чином для класифікації. Алгоритм дерев рішень використовує деревоподібну структуру, де рішення приймаються на основі порогових значень ознак. Модель дерева рішень працює як серія питань "так" або "ні". Наприклад, вона може задавати такі питання: "Чи є розмір файлу аномально великим?" або "Чи має файл доступ до чутливих областей системи?". На основі відповідей дерево слідує шляху і приймає остаточне рішення, наприклад "Шкідливе програмне забезпечення" або "Безпечне", досягаючи кінцевого вузла з висновком (рис. 4).

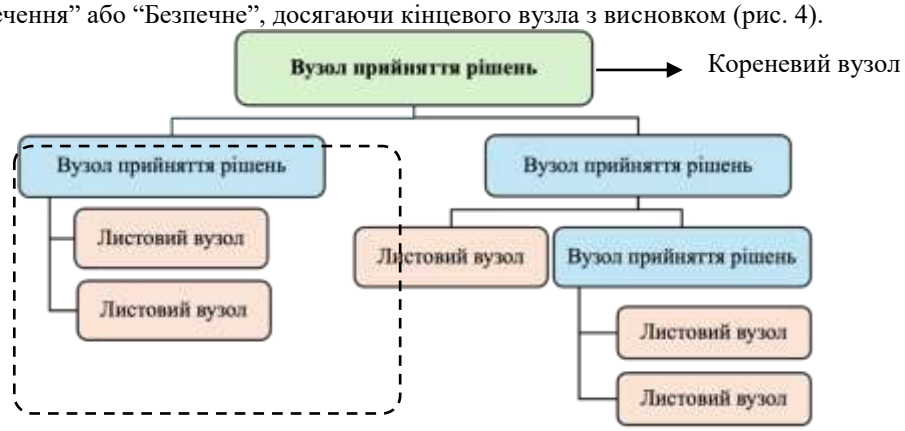

Рис. 4. Візуальне представлення алгоритму дерев рішень

Наївний байєсівський класифікатор (Naive Bayes) використовується в кластеризації та класифікації [20]. Архітектура наївного байєсівського класифікатора базується на умовній ймовірності. Алгоритм створює дерева на основі ймовірності їхнього виникнення. Ці дерева також відомі як байєсівські мережі. Модель наївного байєсівського класифікатора прогнозує щось, обчислюючи ймовірність його виникнення на основі минулих даних. Наприклад, для виявлення спам–листів вона перевіряє, як часто певні ключові слова (наприклад, "безкоштовно" або "виграти") з'являються в спам–листах порівняно зі звичайними. Незважаючи на те, що вона припускає, що кожна ознака є незалежною від інших, вона дуже добре працює для таких завдань, як фільтрування електронної пошти або класифікація тексту (рис. 5).

Рис. 5. Візуальне представлення алгоритму наївного байєсівського класифікатора

Метод опорних векторів (Support Vector Machines) як і два попередніх алгоритми контрольованого МН використовується для класифікації. Метод опорних векторів працює за принципом розрахунку межі. Він проводить межі між класами таким чином, щоб відстань між межею та класами була максимальною, що дозволяє мінімізувати помилки класифікації. Метод розділяє дані на дві групи (наприклад, фішингові веб–сайти та безпечні веб–сайти) шляхом пошуку найкращої межі або лінії (рис. 6). Він аналізує такі особливості, як підозрілі слова в URL–адресі, незвичайні доменні імена, підроблені сертифікати або структуру HTML, і використовує їх для визначення цієї межі. Після навчання метод перевіряє нові веб–сайти і дозволяє вирішити, до якої сторони лінії вони належать – фішингові чи безпечні. У кібербезпеці, крім виявлення фішингових веб–сайтів, вона використовується для виявлення спроб вторгнення в мережевий трафік.

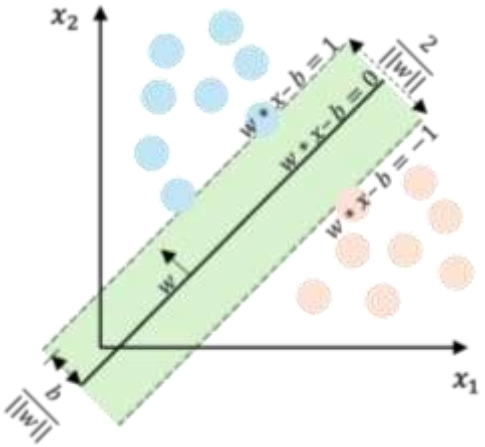

Рис. 6. Візуальне представлення методу опорних векторів

Контрольоване МН широко використовується в системах виявлення вторгнень (IDS) для ідентифікації відомих шаблонів атак та оповіщення команд безпеки при виявленні подібної активності. Однак його ефективність залежить від якості та кількості маркованих даних. Тому така модель має обмеження, коли мова йде про атаки нульового дня.

На відміну від контрольованого навчання, неконтрольоване навчання (Unsupervised Learning) не навчається на маркованих даних. Натомість вони аналізують внутрішню структуру даних для виявлення прихованих шаблонів, аномалій та взаємозв'язків. Ці методи особливо корисні для виявлення аномалій, де метою є виявлення незвичної активності, що відхиляється від звичної поведінки, без попереднього вивчення сигнатур [21]. Методи використовуються для кластеризації та виявлення аномалій, де метою фахівця з безпеки є виявлення закономірностей, які не мають заздалегідь визначених категорій (рис. 7).

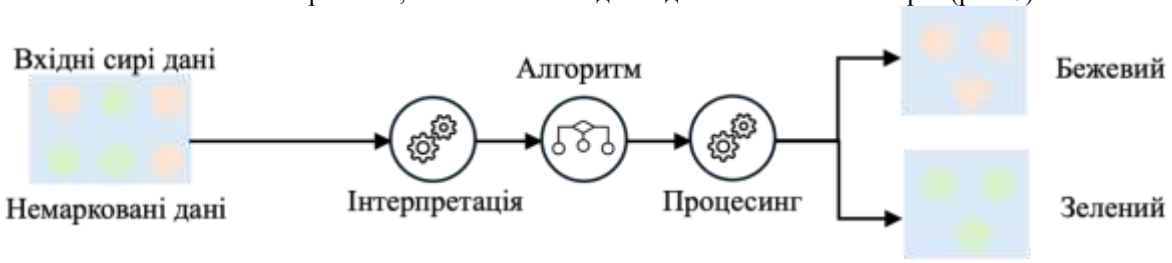

Рис.7. Неконтрольоване МН

Серед усіх алгоритмів неконтрольованого МН в сфері кібербезпеки найчастіше застосовуються кластеризація методом K–середніх (K–Means Clustering), ліс ізоляції (Isolation Forest) та трансформери (Transformers – BERT, GPT).

Кластеризація методом K–середніх – це модель неконтрольованого МН, яка після запуску автоматично створює групи. Елементи, що мають схожі характеристики, розміщуються в одному кластері. Цей алгоритм називається K–середніх, бо він створює K різних кластерів. Середнє значення в конкретному кластері є центром цього кластера [21].

Кластеризація K–Means групує точки даних (наприклад, мережеву активність або записи журналу) у кластери на основі їхньої схожості (рис. 8). Вона працює шляхом пошуку закономірностей у нормальній поведінці, а потім ідентифікації точок даних, які не вписуються в жодну групу, як незвичайні або підозрілі. Це допомагає виявити аномальну активність, наприклад, спробу зловмисника отримати доступ до мережі або несподівану системну помилку.

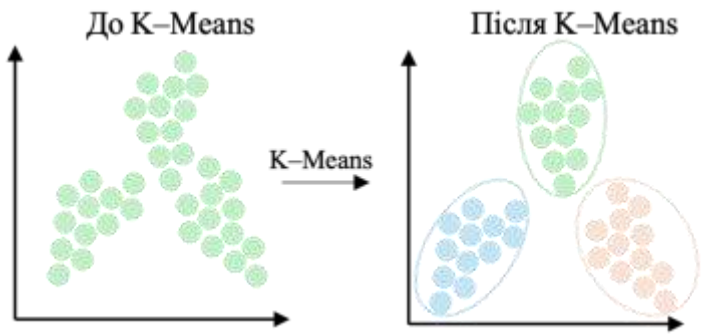

Рис. 8. Візуальне представлення методу K–середніх

У кібербезпеці ця модель МН може використовуватися для виявлення внутрішніх загроз шляхом групування користувачів із подібними паттернами, а також для виявлення відхилень та ідентифікації атак нульового дня.

Ліс ізоляції – модель, що працює як детектив, шукаючи точки даних, що не відповідають іншим. Вона будує багато невеликих дерев рішень і швидко "ізолює" незвичайні дані, оскільки аномальні точки даних потребують менше розділень для їх відокремлення. Це модель виявлення аномалій (рис. 9).

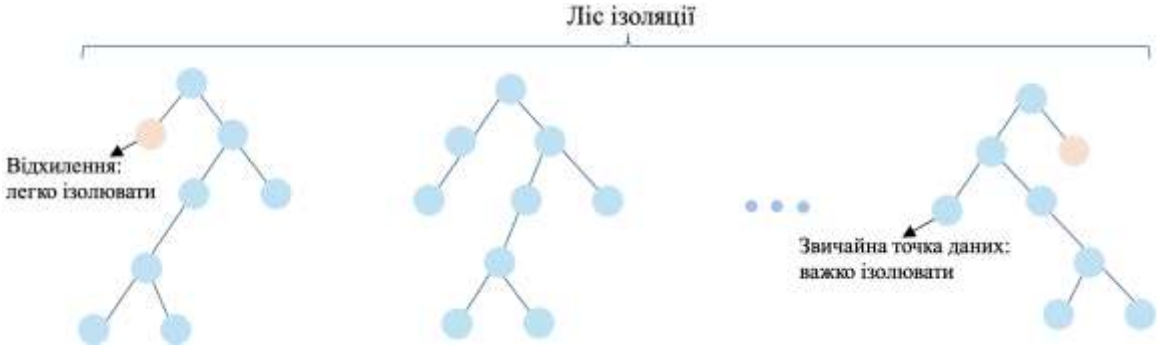

Рис. 9. Візуальне представлення алгоритму ліс ізоляції

У кібербезпеці ця модель може виявляти аномальну поведінку, таку як незвичайні входи в систему або ненормальний мережевий трафік, і використовується в IDS для виявлення аномалій кібератак, а також для аналізу поведінки користувачів, де вона відстежує дії користувачів, щоб виявити відхилення від нормальної поведінки.

Трансформери, такі як BERT (Bidirectional Encoder Representations from Transformers) і GPT (Generative Pre–trained Transformer), є універсальними моделями, які можуть застосовуватися в різних категоріях МН залежно від їхнього використання та методів навчання.

Це потужні комп'ютерні моделі, які розуміють і створюють текст, зосереджуючись на взаємозв'язках між словами в реченні. Вони використовують так звану "увагу", щоб визначити, які слова є найважливішими для значення, навіть у довгих абзацах. Ці моделі навчаються на величезних обсягах тексту з книг, веб–сайтів тощо, що дозволяє їм передбачати, що буде далі в реченні, або точно відповідати на запитання (рис. 10).

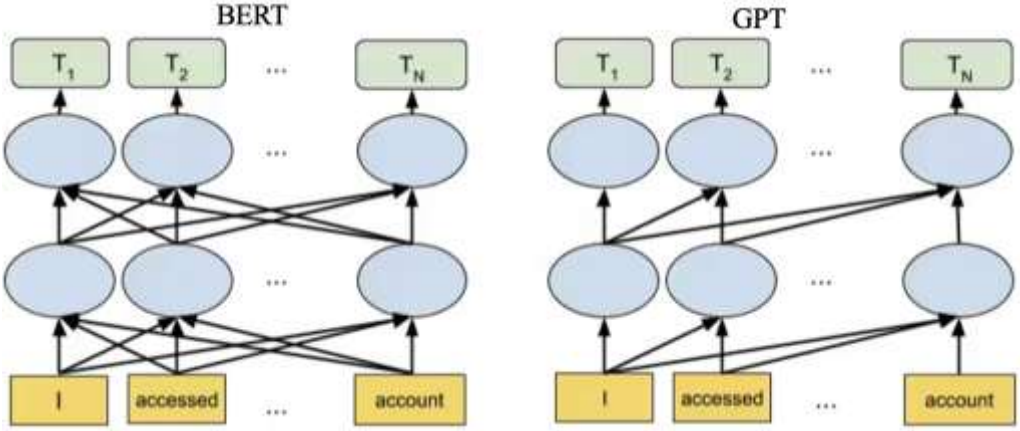

Рис. 10. Візуальне представлення алгоритмів BERT та GPT

Коли такі трансформери працюють у неконтрольованому або самоконтрольованому режимах, вони аналізують великі набори даних без позначених прикладів, щоб виявити закономірності, аномалії та інсайти. Вони особливо цінні в аналізі загроз, виявленні аномалій та розумінні нових технік атак.

Неконтрольоване МН є ефективним методом у дослідницькому аналізі даних, кластеризації та виявленні аномалій, особливо коли марковані дані недоступні. Однак відсутність чітких метрик валідації та залежність від інтерпретації експертів можуть обмежувати його практичне застосування в певних завданнях. Найкращі результати дає баланс між неконтрольованим навчанням та контрольованими техніками.

Навчання з підкріпленням є більш просунутою формою МН, де модель навчається, взаємодіючи зі своїм середовищем і отримуючи зворотний зв'язок у вигляді винагород або штрафів (рис. 11). У кібербезпеці навчання з підкріпленням є особливо перспективним для адаптивних механізмів захисту [22]. Наприклад цей метод може використовуватися в системах запобігання вторгненням (IPS), де система навчається найкращим діям у відповідь на різні типи атак. З часом ця модель оптимізує свої стратегії захисту, покращуючи свою здатність самостійно блокувати або пом'якшувати вторгнення на основі попередніх взаємодій з зловмисниками. Навчання з підкріпленням також застосовується для автоматизованого тестування на проникнення, де системи ШІ досліджують вразливості в мережі та навчаються їх експлуатувати або захищатися від них у контрольованому середовищі.

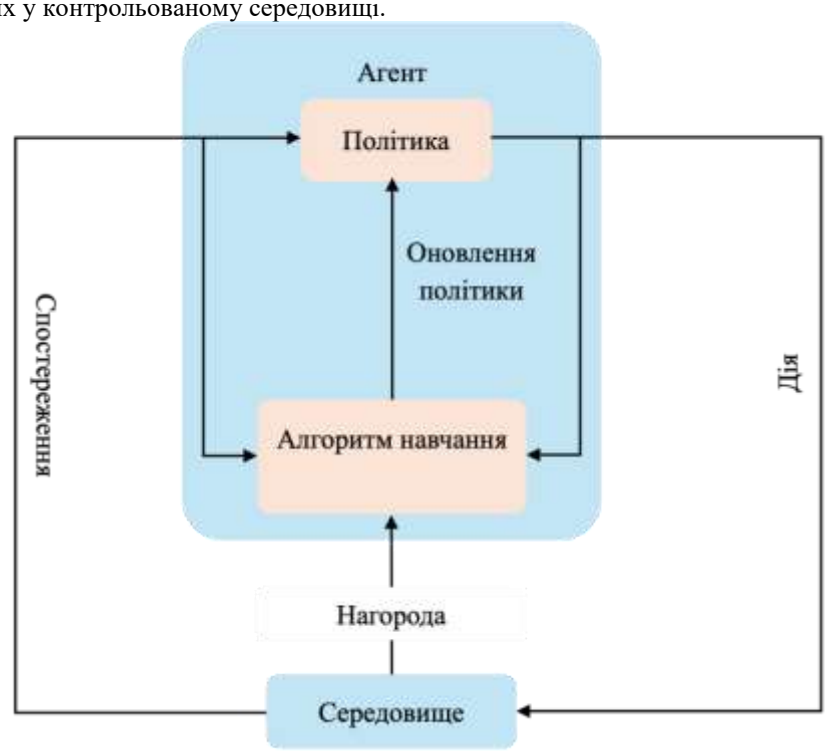

Рис. 11. Навчання з підкріпленням

Глибокі Q–мережі – це алгоритм, який допомагає комп'ютерам приймати рішення, навчаючись на досвіді, як, наприклад, граючи у відеоігри знову і знову, щоб стати кращим гравцем. Глибокі Q–мережі використовують "пам'ять" для зберігання інформації про дії, які добре спрацювали в минулому, і "мозок" (нейронну мережу) для прогнозування найкращих ходів у різних ситуаціях. Випробовуючи різні дії та навчаючись на основі винагород, вона визначає найрозумніший спосіб досягнення своїх цілей, будь то перемога в грі чи зупинка кібератаки. Q–Network базується на Q–Learning [23]. Основною метою Q–Learning є розробка політики, яка направляє агента щодо того, які дії слід вживати в конкретних обставинах, щоб максимізувати винагороди (рис. 12).

В сфері кібербезпеки вона може використовуватися в автоматизованому пошуку загроз шляхом дослідження великих обсягів мережевих журналів або даних про події для виявлення потенційних загроз або аномалій, в IDS шляхом прийняття рішення про блокування або дозвіл з'єднання на основі спостережуваних поведінок, або в моделюванні кібератак та оптимізації захисту шляхом моделювання потенційних сценаріїв атак і навчання ефективним способам реагування для зменшення ризиків.

Рис. 12. Візуальне представлення Q–Learning

Оптимізація проксимальної політики – це алгоритм МН, який вчиться приймати правильні рішення, пробуючи різні варіанти та вдосконалюючись на основі отриманих винагород. Він оновлює свою стратегію, щоб не робити помилок під час навчання, що допомагає залишатися стабільним та ефективним. Оптимізація проксимальної політики часто використовується в роботах, щоб навчити їх діяти в складних ситуаціях, таких як ходьба або гра в шахи (рис. 13).

У кібербезпеці такий алгоритм може навчати системи захищатися від атак, моделюючи різні загрози та навчаючись найкращим реакціям, таким як блокування підозрілого трафіку або ізоляція скомпрометованих пристроїв.

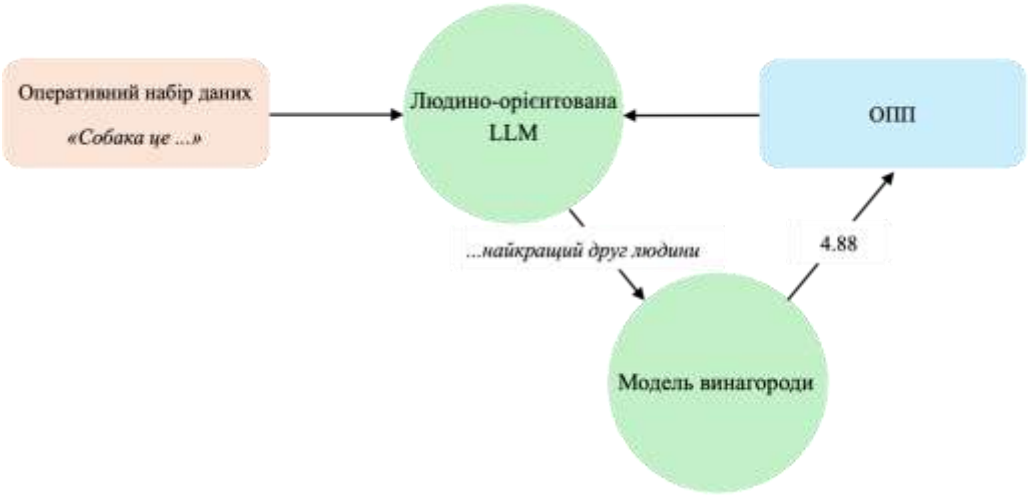

Рис. 13. Візуальне представлення алгоритму оптимізації проксимальної політики

Актор–критик (Soft Actor–Critic) – це алгоритм МН з підкріпленням, що вчиться приймати рішення, балансуючи між двома цілями: отримання найкращих винагород та вивчення різних варіантів. Він намагається не обмежуватися лише безпечними, передбачуваними діями, а й пробує нові речі, щоб знайти ще кращі рішення (рис. 14). Це робить його дуже ефективним в керуванні роботами, безпілотними автомобілями або виявленні аномальних патернів у кібербезпеці, залишаючись при цьому гнучким та ефективним.

Цей алгоритм використовується в IDS, автоматизованому реагуванні на інциденти та захисті кінцевих точок. Наприклад, під час атаки програм–вимагачів актор–критик може вирішити вимкнути уражені системи, мінімізуючи перебої в системі загалом.

Рис. 14. Візуальне представлення алгоритму актор–критик

Розглянуті сучасні методи МН з відповідними алгоритмами забезпечують всебічне охоплення різних аспектів кібербезпеки, зокрема виявлення загроз, реагування на кіберінциденти та моніторинг системи. Кожна модель вирішує специфічні завдання в галузі кіберзахисту, сприяючи створенню надійної та адаптивної системи, здатної ефективно протидіяти новим загрозам. ШІ та МН відіграють ключову роль у розвитку сучасних підходів до кіберзахисту, дозволяючи виявляти аномалії на ранніх етапах, прогнозувати потенційні атаки та автоматизувати процеси реагування.

Для порівняння описаних моделей МН обрано найпоширеніші кількісні метрики, що дозволяють досить точно охарактеризувати роботу моделей з точки зору виявлення кіберінцидентів:

- accuracy (точність) – загальна частка правильно класифікованих прикладів;
- precision (прецизійність) – частка правильно визначених атак серед усіх виявлених;
- recall (повнота) – частка виявлених атак серед усіх дійсних атак;
- F1-score – гармонійне середнє між Precision і Recall;
- False Positive Rate (FPR) – частка хибнопозитивних спрацювань серед нормального трафіку.

Важливо не лише оцінити ефективність та точність виявлення, але й мінімізувати кількість помилкових спрацювань.

Для дослідження використовувався датасет CICIDS 2018, що включає сім різних сценаріїв атак: Brute-force, Heartbleed, Botnet, DoS, DDoS, веб-атаки та проникнення в мережу зсередини. Інфраструктура атаки включає 50 машин, а організація-жертва має 5 відділів і включає 420 машин та 30 серверів. Набір даних включає захоплений мережевий трафік та системні журнали кожної машини, а також 80 функцій, витягнутих із захопленого трафіку за допомогою CICFlowMeter-V3.

Згідно проведених досліджень отримано такі результати:

1. Кероване навчання (Supervised Learning)

Модель	**Accuracy**	**Precision**	**Recall**	**F1-score**	**FPR**
Decision Tree	99.98%	99.97%	99.98%	99.97%	0.01%
Naive Bayes	95.44%	91.87%	94.51%	93.17%	6.12%
Support Vector Machine	97.81%	97.60%	97.94%	97.77%	1.08%

2. Некероване навчання (Unsupervised Learning)

Модель	**Accuracy**	**Precision**	**Recall**	**F1-score**	**FPR**
K-Means Clustering	68.30%	60.02%	74.89%	66.59%	17.65%
Isolation Forest	91.26%	92.77%	88.90%	90.79%	4.62%
BERT (без fine-tuning)	86.42%	83.15%	81.90%	82.52%	7.70%
GPT-2 (без fine-tuning)	84.17%	80.08%	78.52%	79.29%	8.44%

3. Машинне навчання з підкріпленням (Reinforcement Learning)

Модель	**Accuracy**	**Precision**	**Recall**	**F1-score**	**FPR**
Deep Q-Network (DQN)	88.32%	87.12%	86.90%	87.01%	6.95%
PPO	82.47%	80.11%	81.30%	80.70%	9.10%
Soft Actor–Critic (SAC)	84.66%	82.30%	83.19%	82.74%	8.40%

Узагальнююча таблиця

Модель	**Accuracy**	**Precision**	**Recall**	**F1-score**	**FPR**
Decision Tree	99.98%	99.97%	99.98%	99.97%	0.01%
Naive Bayes	95.44%	91.87%	94.51%	93.17%	6.12%
SVM	97.81%	97.60%	97.94%	97.77%	1.08%
K-Means	68.30%	60.02%	74.89%	66.59%	17.65%
Isolation Forest	91.26%	92.77%	88.90%	90.79%	4.62%
BERT	86.42%	83.15%	81.90%	82.52%	7.70%
GPT-2	84.17%	80.08%	78.52%	79.29%	8.44%
DQN	88.32%	87.12%	86.90%	87.01%	6.95%
PPO	82.47%	80.11%	81.30%	80.70%	9.10%
Soft Actor–Critic (SAC)	84.66%	82.30%	83.19%	82.74%	8.40%

Проведене дослідження показало, що серед розглянутих моделей МН – дерев рішень (Decision Tree), наївного байєсівського класифікатора (Naive Bayes), методу опорних векторів (SVM), кластеризації методом K–середніх, ізоляційного лісу (Isolation Forest), трансформерів (BERT, GPT) та алгоритмів з підкріпленням (DQN, PPO, Soft Actor–Critic) – найбільш ефективною, стабільною та практично придатною є модель дерев рішень. Вона показала найвищу точність (99.98%), найменший рівень хибнопозитивних спрацювань (~0.01%), мінімальний час тренування та відсутність потреби в значних обчислювальних ресурсах. Крім того, Decision Tree є інтерпретованою моделлю, що дозволяє фахівцям з кіберзахисту пояснювати та довіряти рішенням системи. Таким чином, модель дерев рішень доцільно рекомендувати як базову для впровадження в системах виявлення кіберінцидентів в ІКС, зокрема в органах, відповідальних за кібербезпеку державного рівня.

Водночас найкращу здатність до виявлення атак нульового дня демонструє алгоритм Isolation Forest. На відміну від моделей, орієнтованих на розпізнавання вже відомих шаблонів атак (наприклад, Decision Tree або SVM), Isolation Forest ґрунтується на принципі виявлення аномалій, тобто виявляє відхилення від нормальної поведінки без потреби у маркованих даних. Це робить його особливо цінним у сценаріях, де нові типи атак ще не були зареєстровані або описані. Отримані кількісні показники (точність понад 91%, прийнятний рівень хибнопозитивних спрацювань) підтверджують практичну придатність цієї моделі для виявлення нових загроз у реальному часі. Таким чином, Isolation Forest є доцільним вибором для систем раннього реагування на кіберінциденти, зокрема для виявлення невідомих або нестандартних атак в умовах сучасних ІКС.

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

1. В результаті дослідження встановлено, сучасні DdoS атаки, фішинг та атаки з використанням ШПЗ, зростають на 25–30% щорічно [6], що підкреслює обмеженість традиційних сигнатурних та поведінкових методів виявлення через їх нездатність ефективно виявляти атаки нульового дня та високий рівень помилкових спрацювань.

2. Показано, що методи МН демонструють значний потенціал у сфері кібербезпеки завдяки здатності до самонавчання та адаптації до нових загроз, що підтверджується даними IBM [16].

3. Подано результати аналізу основних сучасних методі МН.

Встановлено, що контрольовані методи МН забезпечують високу точність для відомих загроз, але обмежені статичністю та залежністю від маркованих даних.

Неконтрольовані методи МН є більш ефективними для виявлення аномалій та невідомих атак, але схильні до помилкових спрацювань через складність інтерпретації результатів; навчання з підкріпленням, перспективне для адаптивного реагування в динамічних середовищах, проте досить ресурсоємне та складне у налаштуванні.

4. Визначені основні недоліки МН, а саме: високий рівень хибних спрацювань, особливо для неконтрольованого навчання, а також брак комплексних досліджень, які б систематизували методи з точки зору мінімізації помилок та пропонували оптимальні рішення для практичного впровадження в ІКС, що вказує на необхідність подальших цільових досліджень.

Література

1. Enisa. (2024, December). 2024 report on the state of cybersecurity in the union (TP–01–24–005–EN–N). European Union Agency for Cybersecurity.

2. National Institute of Standards and Technology. (2024, February). The NIST cybersecurity framework (CSF) 2.0 (NIST CSWP 29). Applied Cybersecurity Division, Information Technology Laboratory, 100 Bureau Drive (Mail Stop 2000), Gaithersburg, MD, USA.

3. ITU Publications. (2024). Global cybersecurity index 2024: 5th edition (ISBN 978–92–61–38751–8). International Telecommunication Union Development Sector.

4. World Economic Forum. (2025, January). Global cybersecurity outlook 2025. Accenture.
5. ISACA. (2024). State of cybersecurity 2024: Global update on workforce efforts, resources, and cyberoperations. Adobe Newsletter.
6. IBM Corporation. (2025, April). IBM X–Force 2025 Threat Intelligence Index (1227cc9e83cb97ae–USEN–07). IBM Institute for Business Value.
7. Державна служба спеціального зв'язку та захисту інформації України. (2025). Російські кібероперації. Аналітика за II півріччя 2024 року.
8. Halbouni, T. S., Gunawan, M. H., Habaebi, M., Halbouni, M., Kartiwi, M., & Ahmad, R. (2022). Machine learning and deep learning approaches for cybersecurity: A review. IEEE Access, 10, 19572–19585. https://doi.org/10.1109/access.2022.3151248
9. Xin, Y., et al. (2018). Machine learning and deep learning methods for cybersecurity. IEEE Access, 6, 35365–35381. https://doi.org/10.1109/access.2018.2836950
10. Bharadiya, J. (2023, June). Machine learning in cybersecurity: Techniques and challenges. European Journal of Technology, 7(2), 1–14. https://doi.org/10.47672/ejt.1486
11. Sindayigaya, L., & Dey, A. (2022, July). Machine learning algorithms: A review. International Journal of Science and Research (IJSR), 11(8), 1127–1133. https://doi.org/10.21275/sr22815163219
12. Gonaygunta, H. (2023, January). Machine learning algorithms for detection of cyber threats using logistic regression. International Journal of Smart Sensor Adhoc Network, 36–42. https://doi.org/10.47893/ijssan.2023.1229
13. Gonaygunta, H. (2023, January). Factors influencing the adoption of machine learning algorithms to detect cyber threats in the banking industry. Faculty of Graduate School, University of the Cumberlands, 96–117.
14. Alshuaibi, M., Almaayah, M., & Ali, A. (2025, February). Machine learning for cybersecurity issues: A systematic review. Journal of Cyber Security and Risk Auditing, 2025(1), 36–46. https://doi.org/10.63180/jcsra.thestap.2025.1.4
15. Mohamed, N. (2025, April). Artificial intelligence and machine learning in cybersecurity: A deep dive into state–of–the–art techniques and future paradigms. Knowledge and Information Systems. https://doi.org/10.1007/s10115–025–02429–y
16. IBM Corporation. (2024, July). Cost of a data breach report 2024.
17. IBM Corporation. (n.d.). What is machine learning (ML)? | IBM. Retrieved July 1, 2025, from https://www.ibm.com/think/topics/machine–learning
18. Солікс Technologies, Inc. (n.d.). ІЩо таке машинне навчання (ML)? | Солікс. Retrieved June 26, 2025, from https://www.solix.com/uk/kb/machine–learning
19. Loza, B. (n.d.). Supervised machine learning in cybersecurity: A comprehensive analysis. Medium. Retrieved June 28, 2025, from https://medium.com/@leev574/supervised–machine–learning–in–cybersecurity–a–comprehensive–analysis–04ac97a822fc
20. Singh, R. (n.d.). Naive Bayes. Medium. Retrieved July 1, 2025, from https://medium.com/@RobuRishabh/naive–bayes–ea63940c5cac
21. Loza, B. (n.d.). Unsupervised machine learning in cybersecurity: A comprehensive analysis. Medium. Retrieved June 22, 2025, from https://medium.com/@leev574/unsupervised–machine–learning–in–cybersecurity–a–comprehensive–analysis–aa4d854e65d2
22. Loza, B. (n.d.). Reinforcement machine learning in cybersecurity: A comprehensive analysis. Medium. Retrieved June 23, 2025, from https://medium.com/@leev574/reinforcement–machine–learning–in–cybersecurity–a–comprehensive–analysis–17c2505c8be0
23. Kovalchuk, G. (n.d.). A beginner's guide to Q–learning: Understanding with a simple gridworld example. Medium. Retrieved June 24, 2025, from https://medium.com/@goldengrisha/a–beginners–guide–to–q–learning–understanding–with–a–simple–gridworld–example–2b6736e7e2c9

References

1. Enisa. (2024, December). 2024 report on the state of cybersecurity in the union (TP–01–24–005–EN–N). European Union Agency for Cybersecurity.
2. National Institute of Standards and Technology. (2024, February). The NIST cybersecurity framework (CSF) 2.0 (NIST CSWP 29). Applied Cybersecurity Division, Information Technology Laboratory, 100 Bureau Drive (Mail Stop 2000), Gaithersburg, MD, USA.
3. ITU Publications. (2024). Global cybersecurity index 2024: 5th edition (ISBN 978–92–61–38751–8). International Telecommunication Union Development Sector.
4. World Economic Forum. (2025, January). Global cybersecurity outlook 2025. Accenture.
5. ISACA. (2024). State of cybersecurity 2024: Global update on workforce efforts, resources, and cyberoperations. Adobe Newsletter.
6. IBM Corporation. (2025, April). IBM X–Force 2025 Threat Intelligence Index (1227cc9e83cb97ae–USEN–07). IBM Institute for Business Value.
7. State Service for Special Communications and Information Protection of Ukraine. (2025). Russian Cyber Operations. Analytics for the Second Half of 2024.
8. Halbouni, T. S., Gunawan, M. H., Habaebi, M., Halbouni, M., Kartiwi, M., & Ahmad, R. (2022). Machine learning and deep learning approaches for cybersecurity: A review. IEEE Access, 10, 19572–19585. https://doi.org/10.1109/access.2022.3151248

9. Xin, Y., et al. (2018). Machine learning and deep learning methods for cybersecurity. IEEE Access, 6, 35365–35381. https://doi.org/10.1109/access.2018.2836950
10. Bharadiya, J. (2023, June). Machine learning in cybersecurity: Techniques and challenges. European Journal of Technology, 7(2), 1–14. https://doi.org/10.47672/ejt.1486
11. Sindayigaya, L., & Dey, A. (2022, July). Machine learning algorithms: A review. International Journal of Science and Research (IJSR), 11(8), 1127–1133. https://doi.org/10.21275/sr22815163219
12. Gonaygunta, H. (2023, January). Machine learning algorithms for detection of cyber threats using logistic regression. International Journal of Smart Sensor Adhoc Network, 36–42. https://doi.org/10.47893/ijssan.2023.1229
13. Gonaygunta, H. (2023, January). Factors influencing the adoption of machine learning algorithms to detect cyber threats in the banking industry. Faculty of Graduate School, University of the Cumberlands, 96–117.
14. Alshuaibi, M., Almaayah, M., & Ali, A. (2025, February). Machine learning for cybersecurity issues: A systematic review. Journal of Cyber Security and Risk Auditing, 2025(1), 36–46. https://doi.org/10.63180/jcsra.thestap.2025.1.4
15. Mohamed, N. (2025, April). Artificial intelligence and machine learning in cybersecurity: A deep dive into state–of–the–art techniques and future paradigms. Knowledge and Information Systems. https://doi.org/10.1007/s10115–025–02429–y
16. IBM Corporation. (2024, July). Cost of a data breach report 2024.
17. IBM Corporation. (n.d.). What is machine learning (ML)? | IBM. Retrieved July 1, 2025, from https://www.ibm.com/think/topics/machine–learning
18. Солікс Technologies, Inc. (n.d.). Що таке машинне навчання (ML)? | Солікс. Retrieved June 26, 2025, from https://www.solix.com/uk/kb/machine–learning
19. Loza, B. (n.d.). Supervised machine learning in cybersecurity: A comprehensive analysis. Medium. Retrieved June 28, 2025, from https://medium.com/@leev574/supervised–machine–learning–in–cybersecurity–a–comprehensive–analysis–04ac97a822fc
20. Singh, R. (n.d.). Naive Bayes. Medium. Retrieved July 1, 2025, from https://medium.com/@RobuRishabh/naive–bayes–ea63940c5cac
21. Loza, B. (n.d.). Unsupervised machine learning in cybersecurity: A comprehensive analysis. Medium. Retrieved June 22, 2025, from https://medium.com/@leev574/unsupervised–machine–learning–in–cybersecurity–a–comprehensive–analysis–aa4d854e65d2
22. Loza, B. (n.d.). Reinforcement machine learning in cybersecurity: A comprehensive analysis. Medium. Retrieved June 23, 2025, from https://medium.com/@leev574/reinforcement–machine–learning–in–cybersecurity–a–comprehensive–analysis–17c2505c8be0
23. Kovalchuk, G. (n.d.). A beginner's guide to Q–learning: Understanding with a simple gridworld example. Medium. Retrieved June 24, 2025, from https://medium.com/@goldengrisha/a–beginners–guide–to–q–learning–understanding–with–a–simple–gridworld–example–2b6736e7e2c9