International Science Journal of Engineering & Agriculture
2025; 4(5): 27-39
https://isg-journal.com/isjea
doi: 10.46299/j.isjea.20250405.03
ISSN: 2720-6319

INTERNATIONAL SCIENCE GROUP

Окремі аспекти проектування системи інформаційної безпеки для захисту IoT-мереж від атак

Ігор Андрущак
Луцький національний технічний університет, Луцьк, Україна
ORCID 0000-0002-8751-4420

Віктор Кошелюк
Луцький національний технічний університет, Луцьк, Україна
ORCID 0000-0002-4136-5087

Анотація: У сучасних умовах стрімкого розвитку Інтернету речей (IoT) питання забезпечення надійної інформаційної безпеки набуває особливого значення. Мережі IoT характеризуються великою кількістю взаємопов'язаних пристроїв, які генерують значні обсяги трафіку та даних, що ускладнює процеси моніторингу та своєчасного виявлення кібератак. Використання традиційних методів захисту виявляється недостатнім, оскільки вони не враховують високий рівень динамічності трафіку та специфіку ресурсно-обмежених пристроїв. У цьому контексті перспективним напрямом є застосування методів машинного навчання для проектування інтелектуальних систем виявлення атак, здатних адаптивно реагувати на нові загрози.

Дослідження зосереджується на аналізі окремих аспектів проектування системи інформаційної безпеки для IoT-мереж, зокрема на використанні алгоритмів Decision Tree (дерево рішень) та K-Nearest Neighbor (KNN). Обидва методи є відомими підходами у сфері класифікації даних, які демонструють високу ефективність у завданнях виявлення вторгнень. Алгоритм Decision Tree дозволяє створити ієрархічну модель прийняття рішень, що забезпечує зрозумілу інтерпретацію результатів та можливість пояснення процесу класифікації. Метод KNN, у свою чергу, базується на аналізі близькості об'єктів у багатовимірному просторі ознак, що дає змогу адаптивно класифікувати нові вхідні дані, спираючись на вже відомі приклади.

Особливу увагу приділено формуванню наборів ознак, що впливають на точність роботи системи. Враховано специфіку IoT-трафіку, а саме: різнорідність протоколів, змінність частоти передачі даних, наявність як звичайних, так і атакувальних шаблонів поведінки. Для експериментів було використано репрезентативні набори даних, що містять приклади як легітимних запитів, так і сучасних поширених атак на IoT-мережі. Результати тестування показали, що обидва алгоритми забезпечують задовільний рівень точності, проте демонструють різні характеристики у контексті швидкодії, чутливості до шуму та вимог до обчислювальних ресурсів.

Наукова новизна роботи полягає в порівняльному аналізі ефективності методів Decision Tree та KNN для завдань виявлення атак у IoT-мережах, а також у визначенні ключових аспектів проектування систем безпеки, орієнтованих на роботу в умовах обмежених ресурсів. Практична значущість полягає у можливості використання отриманих результатів для створення легковагових IDS-систем, придатних для інтеграції у сучасні IoT-середовища.

Таким чином, запропонований підхід дає змогу підвищити рівень захисту IoT-мереж, забезпечивши своєчасне виявлення потенційних атак та зниження ризиків компрометації даних і сервісів.

Ключові слова: IoT-безпека, машинне навчання, кіберзагрози, виявлення аномалій.

1. Вступ

Стрімкий розвиток технологій Інтернету речей (Internet of Things, IoT) створює нові можливості для автоматизації процесів, підвищення ефективності управління та формування інтелектуальних систем у різних сферах – від побуту до промисловості та критичної інфраструктури. Завдяки підключенню великої кількості пристроїв до єдиної мережі виникає можливість обміну даними в режимі реального часу, що відкриває перспективи для створення «розумних» міст, систем охорони здоров'я, виробничих комплексів та транспортних рішень. Водночас таке масштабне зростання кількості підключених пристроїв значно підвищує ризики несанкціонованого доступу, кібератак та порушень конфіденційності даних. Це зумовлює необхідність створення надійних систем інформаційної безпеки, здатних ефективно захищати IoT-мережі від зовнішніх і внутрішніх загроз.

Інтернет речей (IoT, Internet of Things) визначається як ключовий етап еволюції глобального інформаційного простору, що характеризується інтеграцією обчислювальних та комунікаційних можливостей у фізичні об'єкти. У межах цієї концепції матеріальні пристрої набувають здатності до ідентифікації, автономного обміну даними та взаємодії в реальному часі, незалежно від просторово-часових обмежень. IoT репрезентує децентралізовану кіберфізичну інфраструктуру, яка забезпечує інтероперабельність та масштабованість взаємопов'язаних систем, що утворюють основу глобального обчислювального шару [1, 2]. Зазначений шар інтегрує функціональні можливості мережевих протоколів, телекомунікаційних технологій та розподілених цифрових сервісів у єдину архітектуру, що створює умови для автономного функціонування та самоорганізації технічних систем у кіберфізичному середовищі. Стрімке зростання ринку пристроїв Інтернету речей спричинене, в першу чергу, зниженням витрат на їх виробництво та широким доступом до технологій з відкритим кодом. Однак, аналіз, проведений компанією HP у рамках ініціативи OWASP, виявив серйозну проблему: виробники часто не приділяють належної уваги аспектам кібербезпеки. Це робить пристрої IoT надзвичайно вразливими до зловмисних кібервтручань. Крім того, зростання кількості цих пристроїв ускладнює роботу спеціалістів з кібербезпеки, оскільки їм доводиться мати справу з безпрецедентними обсягами даних, що зберігаються та передаються через IoT-мережі [3, 4].

Особливість IoT-мереж полягає у їхній високій гетерогенності та обмежених обчислювальних ресурсах пристроїв. Більшість IoT-девайсів мають обмежений обсяг пам'яті, обчислювальну потужність та енергоресурси, що робить застосування традиційних методів захисту інформації (наприклад, складних криптографічних алгоритмів або ресурсомістких систем виявлення вторгнень) проблематичним. У таких умовах актуальності набувають методи машинного навчання, здатні аналізувати потоки даних у реальному часі та виявляти аномальні дії чи ознаки атак із мінімальним навантаженням на обчислювальні ресурси.

Впровадження ефективних засобів захисту є критично важливим для нейтралізації загроз та мінімізації негативного впливу на IoT-пристрої та користувачів. В умовах постійної еволюції кіберзагроз повне усунення ризиків через моніторинг мережі в реальному часі за допомогою систем виявлення та попередження вторгнень залишається складним завданням. Проте реалізація надійних механізмів контролю доступу та автентифікації здатна значно підвищити рівень захищеності від потенційних атак.

Розповсюдження численних IoT-пристроїв у мережі створює множинні точки вразливості, через які кіберзловмисники та шкідливе програмне забезпечення можуть здійснювати децентралізовані атаки. Впровадження жорстких заходів безпеки призводить до обмеження гнучкості системи та ускладнення архітектури пристроїв, що парадоксально генерує додаткові ризики для кібербезпеки [5]. IoT-екосистема характеризується потребою в диференційованих конфігураціях залежно від специфічних завдань та сфер застосування. Інтегровані механізми захисту мають забезпечувати адаптивність функціонування пристроїв та підтримувати

горизонтальне масштабування для безперешкодного включення нових вузлів до мережевої інфраструктури [6].

Проте переважна частина пристроїв характеризується недостатнім рівнем програмного захисту та містить як виявлені, так і латентні уразливості. Ця обставина створює можливості для кіберзлочинців здійснювати масштабні кібератаки шляхом формування ботнетів на базі пристроїв Інтернету речей [7]. Подібні атаки здатні спричинити руйнування інфраструктурних об'єктів, збої в наданні послуг, значні економічні втрати та підрив іміджу і репутації великих корпорацій. Одним із методів захисту є превентивне блокування подібних ботнетів Інтернету речей через автоматизоване сканування, ідентифікацію вразливих IoT-пристроїв та їх відключення від мережі до моменту компрометації та включення до складу IoT ботнету [8]. Крім того, можна виокремити заходи, спрямовані на детекцію та ізоляцію IoT мереж через точки доступу маршрутизаторів. Такий підхід сприяє попередженню компрометації пристроїв та локалізації IoT ботнетів.

Альтернативний метод передбачає пасивне виявлення та ідентифікацію ботнет-атак шляхом комплексного аналізу мережевого трафіку та внутрішньої активності IoT-пристроїв. Оскільки обробка великих обсягів інформації в режимі реального часу вимагає значних людських ресурсів, це робить процес економічно невигідним. Тому доцільніше застосовувати технології штучного інтелекту та машинного навчання [9].

Штучний інтелект (AI) представляє собою можливість технічних систем відтворювати когнітивні функції людини, тоді як машинне навчання (ML) є підгалуззю AI, що спеціалізується на пошуку, обробці та категоризації інформації. На основі проведеного аналізу система виявляє patterns, формує алгоритми та здійснює автоматичне програмування й налаштування власних компонентів, що може призвести до автономного функціонування [10, 11]. Такий підхід забезпечує оперативне прийняття рішень щодо виявлення кіберзагроз з мінімальною ймовірністю помилок.

Серед різноманіття алгоритмів машинного навчання важливе місце займають методи Decision Tree (дерево рішень) та K-Nearest Neighbor (метод k найближчих сусідів). Їхня привабливість полягає у простоті реалізації, здатності працювати з різними типами даних та відносній інтерпретованості результатів. Алгоритм Decision Tree формує структуру, що дозволяє наочно описати правила класифікації або виявлення вторгнень, що є цінним у завданнях, де пояснюваність рішень має критичне значення. Метод K-Nearest Neighbor, у свою чергу, ефективно працює в умовах, коли необхідно класифікувати нові зразки на основі їх подібності до вже відомих даних. Це дозволяє швидко реагувати на появу нових типів атак, якщо вони мають спільні характеристики з відомими.

Важливим завданням проєктування системи інформаційної безпеки для IoT-мереж є забезпечення балансу між точністю виявлення загроз, швидкістю обробки даних та ефективністю використання обмежених ресурсів. У цьому контексті поєднання класичних методів машинного навчання, таких як Decision Tree та K-Nearest Neighbor, може дати оптимальні результати. Дерево рішень забезпечує структурований підхід до виявлення аномалій, а KNN дозволяє ефективно класифікувати події в режимі реального часу, що створює передумови для формування гнучких та адаптивних систем безпеки.

Крім того, варто зазначити, що сучасні дослідження у сфері захисту IoT-мереж часто базуються на використанні відкритих датасетів, що містять як нормальні сценарії роботи, так і різні типи атак – від сканування портів до складних DDoS-кампаній. Використання таких наборів даних у поєднанні з алгоритмами машинного навчання дає можливість створювати моделі, здатні не лише виявляти загрози, але й адаптуватися до нових векторів атак, які постійно еволюціонують.

З метою реалізації комплексного захисту та імплементації фундаментальних принципів інформаційної безпеки – ідентифікації та автентифікації користувачів, керування правами доступу, забезпечення конфіденційності даних, підтримання їх цілісності, а також механізмів

неспростовності в IoT-середовищі – необхідна модернізація існуючих підходів до забезпечення кібербезпеки [12].

Отже, дослідження окремих аспектів проєктування системи інформаційної безпеки з використанням методів Decision Tree та K-Nearest Neighbor є актуальним науковим і практичним завданням. Воно спрямоване на пошук ефективних підходів до захисту IoT-мереж, що поєднують у собі простоту реалізації, високу точність класифікації та адаптивність до змінного середовища кіберзагроз.

2. Об'єкт і предмет дослідження

Об’єктом дослідження є інформаційна безпека розподілених IoT-мереж, зокрема методи, механізми та архітектурні рішення, спрямовані на забезпечення цілісності, конфіденційності та доступності даних у середовищі з великою кількістю підключених пристроїв. IoT-мережі є складними кіберфізичними системами, які поєднують сенсори, виконавчі пристрої, шлюзи та хмарні сервіси, що створює велику площину потенційних атак. Зростання кількості пристроїв, обмеженість їхніх обчислювальних ресурсів і постійна взаємодія в реальному часі роблять ці системи вразливими до різноманітних атак, включаючи DDoS, підміну даних, несанкціонований доступ і зараження шкідливим ПЗ. Отже, об’єкт дослідження охоплює весь комплекс проблем забезпечення кіберзахисту в IoT-інфраструктурі.

Предметом дослідження є ключові елементи проєктування системи інформаційної безпеки для IoT-мереж із використанням технологій машинного навчання для виявлення та запобігання атакам. Це включає розробку підходів до збору та попередньої обробки телеметричних даних із пристроїв, побудову моделей машинного навчання для класифікації аномалій та виявлення шкідливої активності, а також інтеграцію цих моделей у архітектуру системи захисту. Предметом також є вибір алгоритмів (наприклад, дерева рішень, метод опорних векторів, глибинні нейронні мережі) та оптимізація їх параметрів для досягнення високої точності і швидкодії в умовах обмежених ресурсів. Особлива увага приділяється розробці адаптивних і масштабованих рішень, здатних працювати в гетерогенних і динамічних IoT-середовищах. Таким чином, предмет дослідження охоплює як теоретичні аспекти (моделі та алгоритми машинного навчання), так і практичні питання їх інтеграції в комплексні системи захисту.

3. Мета та задачі дослідження

Метою дослідження є розробка та обґрунтування ключових елементів системи інформаційної безпеки для захисту IoT-мереж від атак із використанням методів машинного навчання, що дозволяють підвищити ефективність виявлення аномалій, швидкість реагування на загрози та загальний рівень кіберзахисту розподілених IoT-систем. Дослідження спрямоване на створення адаптивних, масштабованих і ресурсоефективних рішень для реалістичних сценаріїв використання IoT-пристроїв.

Досягнення поставленої мети передбачало виконання наступних завданнь:

- ✓ аналіз сучасних загроз: провести глибокий аналіз і класифікацію основних видів атак, спрямованих на IoT-пристрої та мережі. Визначити їхні уразливості, механізми експлуатації та потенційні наслідки;
- ✓ огляд існуючих підходів: дослідити та критично оцінити поточні методи захисту IoT-мереж, зокрема традиційні механізми безпеки, виявити їхні сильні та слабкі сторони, особливо щодо адаптації до нових видів атак;
- ✓ вибір та обґрунтування алгоритмів: проаналізувати різні алгоритми машинного навчання (класифікатори, кластеризація, нейронні мережі) та вибрати найбільш підходящі для виявлення аномалій у мережевому трафіку IoT. Обґрунтувати їхню ефективність та здатність до навчання на великих обсягах даних;

✓ проектування архітектури системи: розробити архітектуру системи інформаційної безпеки, яка інтегрує обрані моделі машинного навчання. Спроектувати її компоненти, включаючи модулі збору даних, попередньої обробки, навчання моделі та виявлення аномалій.

✓ створення та тестування моделі: розробити та реалізувати прототип системи. сформувати навчальний датасет, що містить дані як з нормального, так і з аномального трафіку. Провести емпіричне тестування моделі, оцінивши її точність, швидкість відгуку та стійкість до хибних спрацьовувань.

✓ формулювання рекомендацій: на основі отриманих результатів сформулювати практичні рекомендації щодо впровадження та масштабування подібних систем безпеки в реальних IoT-екосистемах.

4. Методи досліджень

Методи машинного навчання, зокрема ті, що базуються на виявленні аномалій, є надзвичайно поширеними в сфері мережевої безпеки. Це пов'язано з їхньою здатністю ефективно ідентифікувати зловмисні дії, що відрізняються від звичайної поведінки мережі. Завдяки класифікації та кластеризації, ці підходи допомагають системам виявлення вторгнень (IDS) ідентифікувати нові, раніше невідомі атаки. Попри свою ефективність, ці технології досі вважаються незрілими для широкого впровадження в комерційних інструментах. Однак, саме застосування машинного навчання для аналізу мережевих даних дозволяє створювати надійніші системи IDS, які можуть виявляти загрози, спираючись на накопичений досвід і розпізнавання аномалій, що є ключовим для боротьби з кібератаками.

Дослідження базується на комплексному підході до аналізу загроз інформаційної безпеки в IoT-мережах із використанням методів машинного навчання для виявлення та класифікації кіберзагроз. Основою методології є порівняльний аналіз ефективності алгоритмів Decision Tree та K-Nearest Neighbor (K-NN) у контексті захисту пристроїв Інтернету речей. Дослідження ключових елементів проєктування системи інформаційної безпеки для захисту IoT-мереж від атак ґрунтується на комбінації теоретичного аналізу, моделювання та машинного навчання. Методологія включає кілька взаємопов’язаних етапів, які дозволяють системно оцінити ефективність захисних механізмів та визначити оптимальні підходи для розпізнавання і нейтралізації атак. На рисунку 1 показано ключові методи машинного навчання, які використовуються в розробці захисних рішень для IoT-мереж.

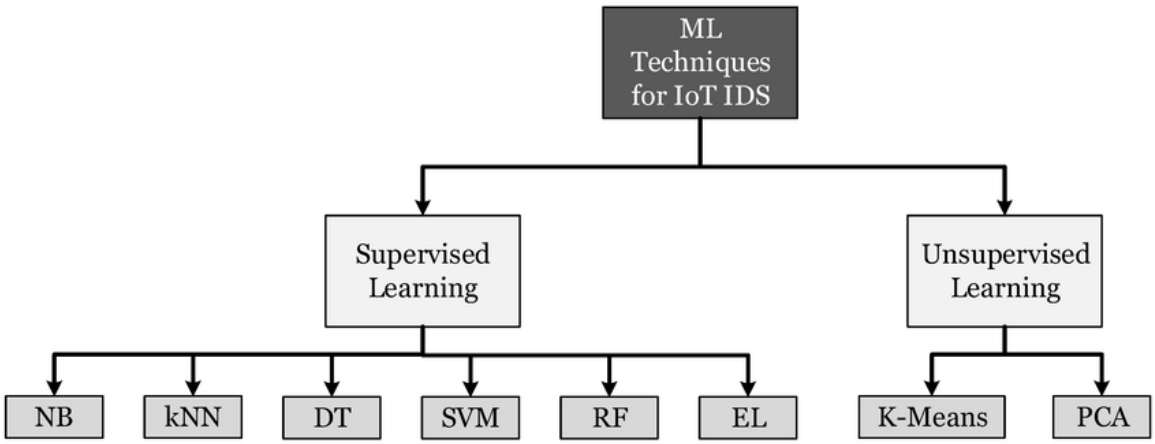

Рис 1. Таксономія методів ML на основі IoT [13].

На початковому етапі здійснюється огляд сучасних досліджень у галузі безпеки IoT-мереж, зокрема робіт, присвячених виявленню атак на рівнях сенсорів, шлюзів та хмарних сервісів. Формалізується задача класифікації мережевих подій як задача машинного навчання, де кожен вхідний запис описує стан мережі у вигляді вектора ознак (кількість пакетів, розподіл

портів, затримка, аномальні патерни). Формується набір атак, що підлягають детекції: DDoS, spoofing, port scanning, sinkhole та інші поширені загрози.

Метод дерева рішень (Decision Tree)

Алгоритм дерева рішень [14] застосовується для класифікації типів атак на IoT-пристрої шляхом побудови ієрархічної структури правил прийняття рішень. Метод базується на принципі рекурсивного поділу набору даних за ознаками, що забезпечують найкращу сепарацію класів загроз. Реалізація в контексті IoT-безпеки передбачає :

- ✓ збір даних: формування датасету з характеристик мережевого трафіку IoT-пристроїв, включаючи параметри з'єднань, частоту запитів, розмір пакетів, протоколи зв'язку;
- ✓ попередня обробка: нормалізація даних, обробка відсутніх значень, кодування категоріальних змінних;
- ✓ побудова моделі: використання критеріїв інформаційного приросту (Information Gain) або індексу Джині для визначення оптимальних точок розбиття;
- ✓ валідація: застосування методу перехресної валідації для оцінки стійкості моделі.

Інтерпретованість результатів дозволяє експертам з кібербезпеки розуміти логіку класифікації загроз, що критично важливо для прийняття обґрунтованих рішень щодо захисту мережі. На основі тренувальних даних будується модель Decision Tree, яка виконує ієрархічну класифікацію подій. Гіперпараметри підбираються експериментально за допомогою GridSearchCV, оптимізуючи глибину дерева, мінімальну кількість зразків у вузлах та критерій розщеплення (gini/entropy). Основна перевага цього підходу полягає у високій інтерпретованості: модель дозволяє виявити найважливіші ознаки, що впливають на визначення атаки.

Метод k-найближчих сусідів (K-Nearest Neighbor)

Для порівняння застосовується метод K-Nearest Neighbor, який відносить нові зразки до класів на основі відстані до найближчих сусідів у просторі ознак. Вибір оптимального параметра *k* здійснюється на основі перехресної валідації. Використовуються різні метрики відстані (евклідова, мангеттенська), щоб визначити, яка з них дає кращу точність для даного типу мережевого трафіку [15].

K-NN алгоритм застосовується для ідентифікації аномального трафіку в IoT-мережах на основі принципу подібності поведінкових патернів. Метод класифікує нові зразки мережевої активності шляхом аналізу k найближчих об'єктів у просторі ознак. Імплементація для IoT-безпеки передбачає :

- ✓ визначення метричної відстані: використання евклідової відстані або метрики Мінковського для обчислення подібності між зразками трафіку;
- ✓ оптимізація параметра k: експериментальне визначення оптимального значення k через аналіз кривих навчання та валідаційних метрик;
- ✓ масштабування ознак: стандартизація параметрів мережевого трафіку для забезпечення рівномірного впливу всіх характеристик;
- ✓ вибір ваг: застосування зваженої схеми голосування, де ближчі сусіди мають більший вплив на класифікацію.

Витягнення цінних даних для аналізу передбачає дослідження пакетів і потоків трафіку в мережах IoT. Цей процес складається з двох основних етапів: спочатку здійснюється захоплення трафіку у вихідному форматі pcap, а потім виконується його екстракція та конвертація у формат CSV для подальшої обробки та аналізу. Для роботи була обрана достовірна збірка даних, яка містить як типовий сучасний мережевий трафік, так і різні сценарії кібератак. Це дозволило отримати повну структуру інформації про трафік, необхідну для подальшого аналізу.

Для зменшення розмірності великого набору ознак і прискорення обробки даних у дослідженні буде використано метод головних компонент (PCA). Основна мета запропонованої нами моделі – досягти максимальної швидкості та точності виявлення вторгнень. Цього буде досягнуто шляхом мінімізації середньоквадратичної похибки та

застосування перехресної перевірки. Для всебічного аналізу ми порівняємо результати роботи двох моделей.

Хоча метод головних компонент (PCA) не є прямою технікою виявлення аномалій, його часто застосовують для зменшення розмірності великих наборів даних. Цей підхід дозволяє перетворити складний набір змінних на меншу, керовану кількість головних компонентів, зберігаючи при цьому основну інформацію. Після того, як PCA зменшив розмірність даних, ці нові, спрощені набори ознак можна ефективно використовувати разом з іншими класифікаторами машинного навчання (ML). Така комбінація є дуже популярною в дослідженнях, особливо для виявлення аномалій у мережах Інтернету речей (IoT). Наприклад, PCA допомагає відфільтрувати надлишок інформації, що дозволяє класифікаторам, таким як нейронні мережі або опорні вектори, точніше і швидше знаходити незвичайну поведінку або збої в мережі.

Для побудови моделі нейронної мережі, призначеної для виявлення атак на мережеві ресурси, у даному дослідженні застосовується підхід навчання з учителем. Процес навчання здійснюється з використанням відкритого набору даних IoTNet24, який широко використовується в академічних і прикладних дослідженнях для оцінювання систем виявлення вторгнень у мережах Інтернету речей.

Набір даних IoTNet24 містить як зразки нормального мережевого трафіку, так і різноманітні типи сучасних атак, що максимально відображають реальні сценарії кіберзагроз. Датасет також включає результати детального аналізу мережевих потоків, зібраних за допомогою інструменту Wireshark, із ретельною розміткою на основі часових міток, IP-адрес джерела та призначення, портів, мережевих протоколів та категорій атак. Усі дані представлені у зручному табличному форматі CSV, що спрощує їх інтеграцію в процес навчання та подальше застосування для побудови й оцінювання моделей машинного навчання.

У якості основного інструментарію для розробки програмного забезпечення в рамках даного дослідження було обрано мову програмування Python. Цей вибір базується на комплексному аналізі технічних вимог проекту та специфіки завдань, пов'язаних з розробкою системи інформаційної безпеки для захисту IoT-мереж. Python володіє найбільш розвиненою екосистемою спеціалізованих бібліотек для роботи з алгоритмами машинного навчання та аналізу даних. Зокрема, для реалізації методів Decision Tree та K-Nearest Neighbor використовуються такі потужні інструменти як Scikit-learn, що надає готові оптимізовані реалізації цих алгоритмів. Додатково, бібліотеки NumPy та Pandas забезпечують ефективну роботу з великими масивами даних та їх обробку. Таким чином, вибір Python як основної мови програмування для даного дослідження є оптимальним рішенням, що забезпечує високу швидкість розробки, надійність реалізації алгоритмів машинного навчання та можливість ефективної обробки великих обсягів даних IoT-мереж.

Проведення дослідження зумовило вибір фундаментальних бібліотек Python серед інструментів для наукових обчислень та аналізу даних. NumPy є незамінним компонентом екосистеми Python для роботи з числовими масивами та математичними операціями високої складності. Основною перевагою NumPy є його оптимізована архітектура для виконання векторних та матричних обчислень. Бібліотека реалізована з використанням мови C, що забезпечує високу швидкість виконання операцій порівняно з чистим Python. Багатовимірні масиви (ndarray) NumPy дозволяють ефективно зберігати та обробляти великі обсяги числових даних з мінімальними витратами пам'яті. Функціональність NumPy включає широкий спектр математичних функцій: від базових арифметичних операцій до складних алгебраїчних перетворень, статистичних розрахунків та операцій з матрицями. Наявність потужного набору вбудованих функцій значно спрощує процес програмування та дозволяє уникнути написання надмірно складного коду для стандартних обчислювальних задач.

Для аналізу та маніпулювання даними в екосистемі Python ми обрали Pandas. Центральними структурами даних Pandas є DataFrame та Series, які забезпечують гнучкі

можливості для зберігання, індексування та обробки табличних даних. DataFrame представляє двовимірну структуру, аналогічну електронній таблиці або реляційній базі даних, в той час як Series являє собою одновимірний масив з мітками індексів. Функціональні можливості Pandas охоплюють широкий спектр операцій з даними: завантаження та експорт даних у різних форматах (CSV, Excel, JSON, SQL, HDF5), очищення та попередня обробка даних, групування та агрегація, злиття та об'єднання наборів даних, а також статистичний аналіз. Бібліотека надає ефективні алгоритми для роботи з відсутніми значеннями, дублікатами та аномаліями в даних.

Matplotlib у поєднанні з Seaborn формують потужну екосистему для створення високоякісних візуалізацій даних у Python. Matplotlib служить базовою платформою, що надає низькорівневий контроль над кожним елементом графічного відображення, від осей координат до кольорових схем. Цей інструментарій дозволяє дослідникам та аналітикам створювати широкий спектр візуальних представлень: від простих лінійних графіків та гістограм до складних багатовимірних візуалізацій та інтерактивних діаграм. Seaborn, як високорівнева надбудова над Matplotlib, значно спрощує процес створення статистично інформативних візуалізацій. Бібліотека пропонує готові шаблони для побудови теплових карт, розподілів, кореляційних матриць та регресійних моделей, автоматично застосовуючи естетично привабливі стилі та кольорові палітри. Особливою перевагою Seaborn є її інтеграція з pandas DataFrame, що дозволяє безпосередньо працювати зі структурованими даними без додаткових перетворень. Комбінування цих двох бібліотек забезпечує оптимальний баланс між гнучкістю налаштувань та простотою використання, що робить їх незамінними інструментами для exploratory data analysis (EDA) та презентації результатів дослідження.

Scikit-learn являє собою найбільш всеосяжну та зріла бібліотеку машинного навчання в екосистемі Python, що охоплює практично всі аспекти сучасних алгоритмів ML. Архітектура бібліотеки побудована навколо універсального API, який забезпечує консистентний підхід до роботи з різними типами алгоритмів: від простих лінійних моделей до складних ансамблевих методів.

Функціональність Scikit-learn включає повний спектр задач машинного навчання:

- ✓ Supervised learning: класифікація та регресія з використанням алгоритмів від логістичної регресії до gradient boosting
- ✓ Unsupervised learning: кластеризація, зменшення розмірності, виявлення аномалій
- ✓ Preprocessing та feature engineering: нормалізація, енкодинг категоріальних змінних, селекція ознак
- ✓ Model selection та evaluation: крос-валідація, гіперпараметрична оптимізація, метрики якості

Особливою цінністю Scikit-learn є її бездоганна документація, що містить не лише технічні специфікації API, але й детальні пояснення математичних основ алгоритмів, практичні приклади та рекомендації щодо вибору моделей для конкретних задач. Велика колекція tutorial’ів та прикладів використання робить бібліотеку доступною як для досвідчених data scientist’ів, так і для новачків у сфері машинного навчання.

Завдяки активній спільноті розробників та регулярним оновленням, Scikit-learn залишається золотим стандартом для реалізації класичних алгоритмів машинного навчання, забезпечуючи надійну основу для розв'язання широкого спектра аналітичних задач.

5. Результати досліджень

Навчання моделі відбуватиметься на датасеті IoTNet24 [16]. Цей датасет представлений у форматі CSV-файлу, де записи розділені комами, а кожен новий рядок відповідає окремому набору даних. Файл можна переглянути за допомогою будь-якого текстового редактора. Файл містить понад 23 000 записів (рядків) та шість ключових характеристик, отриманих в результаті обробки за допомогою Zeek, виконаної творцями набору даних. Ці записи були отримані з мережевого трафіку IoT-пристроїв. Датасет IoTNet24 Dataset for IDS містить шість

ключових ознак, отриманих з обробки мережевого трафіку: duration (розрізняє короткі безпечні сесії та довгі сесії атак (наприклад, DoS, груба сила); proto (критично важливо для виявлення атак, специфічних для протоколу, та класифікації типів потоків); service (допомагає відокремити безпечний трафік від атак, спрямованих на конкретні сервіси (наприклад, тунелювання DNS, груба сила SSH); src_bytes (корисно для виявлення витоків даних або аномальних шаблонів завантаження); dst_bytes (допомагає виявляти такі сценарії, як атаки відбиття або аномальні корисні навантаження відповіді); conn_state (вказує, чи сеанс завершено, скинуто чи закінчився невдачею – сильний сигнал для сканування та невдалих атак на з'єднання).

Спираючись на огляд літератури та ефективність використання методів машинного навчання для захисту та розпізнавання атак на інтернет речей в роботі використовується дерево прийняття рішень (Decision Tree, DT) та K-найближчий сусід (K-Nearest Neighbor, KNN). З метою підвищення точності та узагальнювальної здатності моделей здійснюється оптимізація їхніх гіперпараметрів. Крім того, поширеною практикою є застосування методу головних компонент (PCA, Principal Component Analysis) для попередньої трансформації та зменшення розмірності даних перед їх поданням на вхід моделі.

Для експерименту використаємо модель RandomForestClassifier з модуля sklearn.ensemble, налаштовану лише з параметром max_depth=2, та проведемо оцінювання її валідаційних характеристик:

✓ середньоквадратична похибка (Mean squared error, MSE) передбачення для тренувального набору даних. Позначимо цей параметр як MSE_TRAIN;

✓ перехресна перевірка (Cross-validation, CV) тренувального набору (cv=3, scoring='neg_mean_squared_error'). Позначимо цей параметр як CV_SCORE.

✓ середньоквадратична похибка (Mean squared error, MSE) передбачення для тестового набору даних. Позначимо цей параметр як MSE_TEST.

На основі аналізу IoTNet24 Dataset for IDS визначимо головні ознаки базової моделі, що продемонстровано на рисунку 2.

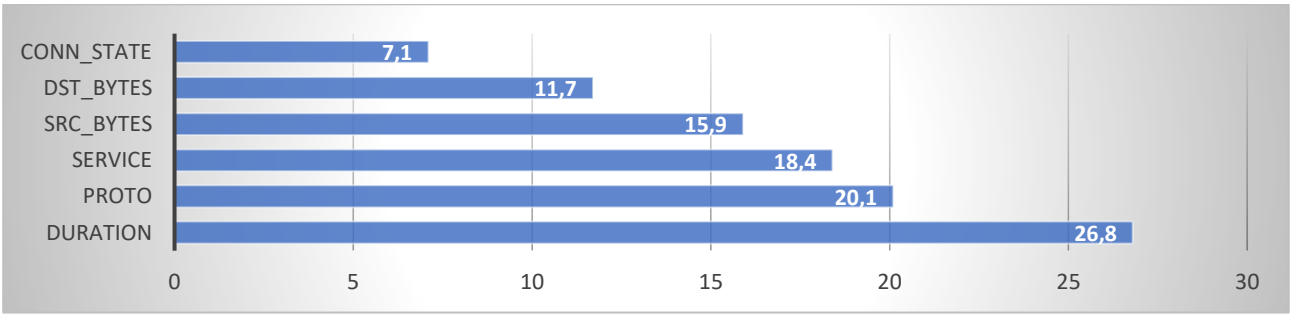

Рис 2. Головні ознаки базової моделі (авторська розробка).

Набір даних спочатку буде розділено на тренувальний (90%) і тестовий (10%) за допомогою функції train_test_split з бібліотеки Scikit-learn. Таке співвідношення є доцільним завдяки великому об'єму даних. Перед початком роботи необхідно масштабувати дані, оскільки їхні значення представлені в різних одиницях вимірювання. Це важливо для коректної роботи алгоритмів. Для виконання стандартизації даних ми використаємо інструмент StandardScaler з бібліотеки sklearn.preprocessing. Аналіз валідаційних показників моделей, які включають середньоквадратичну похибку на тренувальних і тестових даних, а також результати крос-валідації, дозволяє зробити висновки, представлені в Таблиці 1.

Таблиця 1. Зведена таблиця валідаційних характеристик

Модель	Критерій валідації	Дерева прийняття рішень (Decision Tree, DT)	K-Найближчий сусід (K-Nearest Neighbor, KNN)
Базова модель	MSE_TRAIN	0.008	0.003
	CV_SCORE	-0.008432	-0.003194
	MSE_TEST	0.007	0.001
Базова модель з методом головних компонентів (PCA)	MSE_TRAIN	0.011	0.005
	CV_SCORE	-0.010274	-0.00461
	MSE_TEST	0.010	0.004
Базова модель з PCA та Randomized Search CV	MSE_TRAIN	0.012	*
	CV_SCORE	-0.013918	*
	MSE_TEST	0.011	*

Застосування зниження розмірності за допомогою методу головних компонентів та вичерпна оптимізація гіперпараметрів не гарантують підвищення продуктивності моделі порівняно з її базовою версією, яка використовує стандартні налаштування.

6. Висновки

Технологія Інтернету речей (IoT) об'єднує різноманітні “розумні” пристрої через глобальну мережу. Однак, традиційні методи захисту виявляються недостатніми для протидії численним кібератакам на ці пристрої. Для підвищення рівня безпеки та доповнення існуючих систем захисту, все частіше використовується система виявлення вторгнень (IDS). Застосування цієї системи дозволяє аналізувати мережевий трафік, виявляти ознаки атак та оперативно реагувати на них, що допомагає мінімізувати потенційну шкоду.

У ході дослідження було розглянуто окремі аспекти проєктування системи інформаційної безпеки, орієнтованої на захист мереж Інтернету речей (IoT) від атак різної природи. Особливу увагу було приділено методам машинного навчання Decision Tree (DT) та K-Nearest Neighbor (KNN), які продемонстрували свою ефективність у вирішенні завдань виявлення аномалій та класифікації мережевого трафіку. Вибір саме цих методів зумовлений тим, що вони поєднують відносну простоту реалізації з високою продуктивністю у випадках, коли необхідно швидко обробляти великі масиви даних, характерні для IoT-середовищ.

Метод Decision Tree дав змогу продемонструвати високу інтерпретованість результатів, що є важливою перевагою для адміністраторів безпеки. Завдяки ієрархічній структурі дерева прийняття рішень забезпечується зрозуміле пояснення процесу класифікації, що підвищує довіру до роботи системи виявлення атак. Крім того, DT добре працює з різнорідними даними, здатний обробляти як числові, так і категоріальні ознаки, що актуально у випадку IoT-трафіку, де спостерігається велика варіативність параметрів. Водночас до недоліків цього підходу належить схильність до переобучення, особливо при великій кількості ознак, що потребує ретельної оптимізації глибини дерева та використання додаткових методів (наприклад, ансамблевих).

Метод K-Nearest Neighbor зарекомендував себе як універсальний інструмент для задач класифікації з використанням метрик відстані. Його застосування у сфері безпеки IoT-мереж дозволяє досягати високої точності виявлення відомих та схожих за характеристиками атак, а також гнучко налаштовувати алгоритм залежно від розміру вибірки чи кількості сусідів (k). Суттєвою перевагою KNN є відсутність складного етапу навчання — алгоритм орієнтований на використання збережених зразків, що робить його придатним для динамічних середовищ з часто оновлюваними даними. Проте недоліком методу є його обчислювальна складність під час класифікації, особливо за умов великих IoT-мереж, а також чутливість до вибору метрики відстані та нормалізації даних.

Порівняльний аналіз продемонстрував, що обидва методи є релевантними для використання у системах інформаційної безпеки IoT, проте їхня ефективність суттєво залежить від контексту застосування. У випадку, коли ключовим критерієм є зрозумілість і

швидкість ухвалення рішень, доцільніше використовувати Decision Tree. Якщо ж основним завданням виступає гнучка адаптація до різноманітних типів атак та швидке оновлення моделей на основі нових даних, метод KNN може виявитися більш придатним.

Загалом проведене дослідження підтвердило важливість інтеграції методів машинного навчання у процес проєктування систем захисту для IoT-мереж. Їх використання дозволяє значно підвищити рівень виявлення атак, скоротити кількість помилкових спрацювань і забезпечити адаптивність до нових загроз. У перспективі найбільш ефективним є комбіноване застосування різних алгоритмів, зокрема використання ансамблевих методів на основі дерев рішень або гібридних моделей, де KNN поєднується з іншими алгоритмами класифікації. Це відкриває можливість створення комплексних рішень, які здатні одночасно задовольняти вимоги до точності, продуктивності та інтерпретованості результатів.

Таким чином, розглянуті підходи формують підґрунтя для подальших досліджень у сфері безпеки IoT-мереж. Їх практична цінність полягає не лише в теоретичній можливості побудови ефективних IDS-систем, але й у потенціалі їхнього впровадження у реальні інфраструктури Інтернету речей, де загрози мають динамічний і багатовекторний характер. Інтеграція цих підходів у єдину систему інформаційної безпеки створює надійний захист IoT-інфраструктури, забезпечуючи баланс між швидкодією та точністю виявлення загроз. Подальші дослідження мають зосередитися на оптимізації параметрів алгоритмів та розробці гібридних моделей для підвищення загальної ефективності системи захисту.

Список літератури:

1) Kumar, Satendra & Kanchan, & Kumar, Amit & Aggarwal, Prachi. (2023). Internet of things (iot) applications and challenges: a review. 11. 359-367.

2) Hornos, M.J., Quinde, M. Development methodologies for IoT-based systems: challenges and research directions. J Reliable Intell Environ 10, 215–244 (2024). https://doi.org/10.1007/s40860-024-00229-9

3) Lan Luo, Christopher Morales-Gonzalez, Shan Wang, Zhen Ling, Xinwen Fu. A Unified View of IoT And CPS Security and Privacy. 2024 International Conference on Computing, Networking and Communications (ICNC). https://doi.org/10.1109/ICNC59896.2024.10555966

4) Ameyed, D., Jaafar, F., Petrillo, . et al. Quality and Security Frameworks for IoT-Architecture Models Evaluation. SN COMPUT. SCI. 4, 394 (2023). https://doi.org/10.1007/s42979-023-01815-z

5) Almutairi, M., & Sheldon, F. T. (2025). IoT–Cloud Integration Security: A Survey of Challenges, Solutions, and Directions. Electronics, 14(7), 1394. https://doi.org/10.3390/electronics14071394

6) Tabi Fouda, B. M., Wang, L., Han, D., Ngoumou, P. C., & Atangana, J. (2025). Design and Implementation of a Novel IoT Architecture for Data Release System Between Multiple Platforms: Case of Smart Offshores. Sensors, 25(11), 3384. https://doi.org/10.3390/s25113384

7) R. Kumar Yadav and K. Karamveer, A Survey on IOT Botnets and their Detection Approaches. 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 1901-1906, doi: 10.1109/ICAC3N56670.2022.10074482.

8) Gelgi, M., Guan, Y., Arunachala, S., Samba Siva Rao, M., & Dragoni, N. (2024). Systematic Literature Review of IoT Botnet DDOS Attacks and Evaluation of Detection Techniques. Sensors, 24(11), 3571. https://doi.org/10.3390/s24113571

9) W, Regis & G, Kirubavathi & K., Sridevi. (2023). Detection of IoT Botnet using Machine learning and Deep Learning Techniques. 10.21203/rs.3.rs-2630988/v1.

10) Sen, R., Heim, G., & Zhu, Q. (2022). Artificial Intelligence and Machine Learning in Cybersecurity: Applications, Challenges, and Opportunities for MIS Academics. Communications of the Association for Information Systems, 51, pp-pp. https://doi.org/10.17705/1CAIS.05109

11) V. Khatri, G. Agarwal, A. K. Gupta and A. Sanghi, Machine Learning and Artificial Intelligence in Cybersecurity: Innovations and Challenges. 2024 Second International Conference on Advanced Computing & Communication Technologies (ICACCTech), Sonipat, India, 2024, pp. 732-737, doi: 10.1109/ICACCTech65084.2024.00122.

12) Mohamed نشأت عبد اللطيف د., Nachaat. (2025). Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms. Knowledge and Information Systems. 67. 6969-7055. 10.1007/s10115-025-02429-y.

13) Ashraf, Javed & Moustafa, Nour & Khurshid, Hasnat & Debie, Essam & Haider, Waqas & Wahab, Abdul. (2020). A Review of Intrusion Detection Systems Using Machine and Deep Learning in Internet of Things: Challenges, Solutions and Future Directions. Electronics. 9. 10.3390/electronics9071177.

14) Decision Tree in Machine Learning. Available at: https://itwiki.dev/data-science/ml-reference/ml-glossary/decision-tree-in-machine-learning

15) K-Nearest Neighbor(KNN) Algorithm. Available at: https://www.geeksforgeeks.org/machine-learning/k-nearest-neighbours/

16) IoTNet24 Dataset for IDS. Available at: https://www.kaggle.com/datasets/wittigenz/hydras

Specific aspects of designing an information security system to protect IoT networks from attacks

Igor Andrushchak
Department of Software Engineering, Lutsk National Technical University, Lutsk, Ukraine
ORCID 0000-0002-8751-4420

Viktor Kosheliuk
Department of Computer Science, Lutsk National Technical University, Lutsk, Ukraine
ORCID 0000-0002-4136-5087

Abstract: In today's rapidly developing Internet of Things (IoT), the issue of ensuring reliable information security is of particular importance. IoT networks are characterized by many interconnected devices that generate significant amounts of traffic and data, which complicates the processes of monitoring and timely detection of cyberattacks. Traditional protection methods are insufficient, since they do not consider the high level of traffic dynamics and the specifics of resource-limited devices. In this context, a promising direction is the application of machine learning methods for designing intelligent attack detection systems that can respond adaptively to new threats.

The study focuses on analyzing individual aspects of designing an information security system for IoT networks, particularly the use of the Decision Tree and K-Nearest Neighbor (KNN) algorithms. Both methods are well-known approaches in data classification, demonstrating high efficiency in intrusion detection tasks. The Decision Tree algorithm allows you to create a hierarchical decision-making model that provides a clear interpretation of the results and the ability to explain the classification process. The KNN method, in turn, is based on the analysis of the proximity of objects in a multidimensional feature space, which allows you to adaptively classify new input data, relying on already known examples.

Special attention is paid to forming feature sets that affect the system's accuracy. The specifics of IoT traffic are considered: the protocols' heterogeneity, the data transmission frequency variability, and the presence of standard and attack behavior patterns. Representative data sets containing

examples of both legitimate requests and modern common attacks on IoT networks were used for experiments. The test results showed that both algorithms provide a satisfactory level of accuracy, but demonstrate different characteristics in the context of speed, noise sensitivity, and computing resource requirements.

The scientific novelty of the work lies in the comparative analysis of the effectiveness of the Decision Tree and KNN methods for detecting attacks in IoT networks and in identifying key aspects of designing security systems focused on working in conditions of limited resources. The practical significance lies in the possibility of using the obtained results to create lightweight IDS systems suitable for integration into modern IoT environments.

Thus, the proposed approach increases the protection level of IoT networks, ensuring timely detection of potential attacks and reducing the risks of compromising data and services.

Keywords: IoT security, machine learning, cyber threats, anomaly detection.