УДК 004.056 DOI 10.36910/775.24153966.2022.74.5

**Є. Б. Козак**
*магістр в галузі комп'ютерних наук, розробникпрограмного забезпечення, інженер-програміст GAN Inc.*

## ТРАДИЦІЙНІ МОДЕЛІ МАШИННОГО НАВЧАННЯ У СФЕРІ ІНТЕРНЕТУ РЕЧЕЙ

***У статті досліджено традиційні моделі машинного навчання у сфері Інтернету речей. Визначено напрямки розповсюдження розумних систем та проаналізовано перспективи розвитку. Описано предмети повсякденного життя які відносяться до групи Інтернету речей. Наголошено, що величезний масштаб мереж IoT приносить нові завдання, такі як управління цими пристроями, величезний обсяг даних, зберігання, зв'язок, обчислення, безпека та конфіденційність. Наріжним каменем комерціалізації технологій IoT є гарантія безпеки та конфіденційності, а також задоволення споживачів. До головних перешкод для підвищення безпеки розумних пристроїв віднесено конкуренцію на ринку та технічні обмеження. Підкреслено, що традиційні моделі машинного навчання спрямовані на характеристику та визначення рівня шкідливих дій пристроїв IoT, навчання та тестування нейронної мережі для класифікації пристроїв IoT на основі особливостей мережевого трафіку для забезпечення класифікації IP-адрес, близьких до реального часу та оцінку ефективності алгоритмів. Проведено класифікацію традиційних алгоритмів машинного навчання: алгоритми навчання з вчителем, без вчителя та підкріплення. Описано кожен з них. Визначено принципи глибокого навчання та навчання з підкріпленням, а також їх поєднання. Наголошено, що традиційні методи машинного навчання, що використовуються в безпеці IoT працюють з маркованими даними і використовуються в мережах IoT для зондування спектру, оцінки каналів, адаптивного фільтрування, проблем безпеки та локалізації. Зазначається, що трафік IoT зазвичай характеризується великим обсягом, різноманітністю, змінною швидкістю та невизначеністю. Сформовано перелік загальних обмежень використання методів машинного навчання у мережах IoT. Підкреслено, що у реальному світі, коли дані з різних джерел мають різне форматування та подання принцип константи первинного набору даних не працює і машинне навчання потребує попередньої обробки та очищення даних перед тим, як помістити їх у певну модель.***

***Ключові слова:*** *Інтернет речей, машинне навчання, модель, кібератака, небезпека, продуктивність.*

**E. Cozac**

## TRADITIONAL MODELS OF MACHINE LEARNING IN THE FIELD OF THE INTERNET OF THINGS

***The article investigates traditional models of machine learning in the field of the Internet of Things. The directions of distribution of intelligent systems are determined and the prospects of development are analyzed. Describes the objects of everyday life that belong to the group of the Internet of Things. It is emphasized that the huge scale of IoT networks brings new tasks, such as managing these devices, huge amounts of data, storage, communication, computing, security and privacy. The cornerstone of the commercialization of IoT technologies is the guarantee of security and confidentiality, as well as consumer satisfaction. The main obstacles to improving the safety of smart devices include market competition and technical constraints. It is emphasized that traditional machine learning models are aimed at characterizing and determining the level of harmful effects of IoT devices, training and testing of the neural network to classify IoT devices based on network traffic to provide classification of IP addresses, close to real time and evaluate algorithms. The classification of traditional machine learning algorithms is carried out: learning algorithms with teacher, without teacher and reinforcement. Each of them is described. The principles of deep learning and reinforced learning, as well as their combination are defined. It is emphasized that in the real world, when data from different sources have different formatting and representation, the constant principle of the primary data set does not work and machine learning requires pre-processing and purification of data before placing them in a particular model.***

***Key words****: Internet of Things, machine learning, model, cyberattack, danger, productivity.*

**Постановка проблеми.** Інтернет речей (IoT) зробив революцію в способах взаємодії з навколишнім середовищем. Розумні автомобілі, розумні міста, розумні будинки тепер реалізуються за допомогою різних вбудованих пристроїв, які працюють практично без втручання людини. Однак ці вбудовані пристрої породжують безліч проблем з безпекою, оскільки більшість виробників, як і раніше, надають більшого значення трьом принципам (прототипування, виробництво і продуктивність), ніж безпеці. Ця вроджена вада проявляється у вигляді різних атак типу «відмова в обслуговуванні» (DoS), організованих за допомогою не запитуваних пристроїв Інтернету речей в Інтернеті. Велика пропускна здатність без необхідності в додаткових ресурсах, що впливає на великомасштабні інфраструктури в Інтернеті. Таким чином, розуміння природи цих атак і швидке виявлення заражених пристроїв стає необхідним для боротьби з цією ситуацією. На сьогодні, необхідним є розгляд традиційних моделей для класифікації не запитуваних пристроїв Інтернету речей на підприємствах, що використовують машинне навчання, а саме інформація заголовка IP з даних даркнета збирається для аналізу.

**Аналіз останніх досліджень і публікацій.** Сучасні наукові здобутки у сфері Інтернету речей охоплюють методи прототипування виробів, які є основою кожної окремої системи,

*© Є. Б. Козак*

виробництво датчиків для застосування у сфері Інтернету речей, продуктивності систем та їх застосування на всіх рівнях реалізації, обходячи питання безпеки у своїй більшості.

Так у [1] розкриваються принципи побудови захищеного інтелектуального простору Інтернету речей. Наведено класифікацію IoT за Робом Ван Краненбургом. Запропоновано розширений алгоритм крихітного шифрування (Extended Tiny Encryption Algorithm, XTEA) – один з найшвидших і найефективніших криптографічних алгоритмів, що існують.

До питання цифрових технологій в умовах формування цифрової економіки детально підійшов А. Ю. Семеног [2]. Автор здійснив аналіз основних технологій, що застосовуються в умовах формування цифрової економіки, визначив їх сутність, види та способи практичного застосування. Дослідив властивості, потенційні переваги та ризики блокчейн-технології, навів приклади компаній, що його використовують. А головне визначив основні елементи та ієрархію інтернету речей. О. Забігайло, О. Терешко, І. Панасюк та Р. Леськів [3] запропонували дослідження механізмів розгортання елементів «розумного міста».

І. Р. Опірський, Р. В. Головчак, І. Р. Мойсійчук, Т. С. Балянда та С. П. Гаранюк [4] визначили та проаналізовали найпоширеніші загрози, з якими може зустрітися користувач при встановленні IoT пристроїв. Науковцями встановлено також, що зазвичай не лише виробник створює загрози безпеці IoT пристроїв. Також наведено ряд порад для користувачів, котрі хочуть знизити ризик витоку даних, пов’язаний із вразливостями систем інтернету речей. Авторами підкреслено, що неодиничними випадками є некоректне налаштування, використання та зберігання таких приладів.

Сучасні тенденції розвитку інформаційних технологій з відокремленням Інтернету речей та можливістю застосування машинного навчання дослідили Д.Г. Косяков та О.В. Ларченко [5].

Із зарубіжних авторів варто відзначити такі роботи як: Yeboah-Ofori, A., Islam, S., Yeboah-Boateng, E. [6], Smith, A. [7], Razali, M. F., Razali, M. N., Mansor, F. Z., Muruti, G., & Jamil, N. [8], Schroer A. [9], Rishi R., Saluja R. [10], O’Halloran D., D’Souza F. [11] та інші.

Проте, враховуючи описані наукові набутки, за темою, питання систематизації та узагальнення традиційних моделей машинного навчання у сфері Інтернету речей залишається відкритим та потребує детального опрацювання.

**Постановка завдання.** У роботі необхідно дослідити традиційні моделі машинного навчання у сфері Інтернету речей.

**Викладення основного матеріалу дослідження.** За останні кілька років на ринку спостерігається масове розповсюдження пристроїв IoT, і передбачається, що ця кількість наблизиться до майже 30 мільярдів пристроїв до 2022 року, що у грошовому еквіваленті складе 53 мільярдів доларів[2]. IoT розглядається як взаємопов'язана та розподілена мережа вбудованих систем, що здійснюють зв'язок за допомогою дротових або бездротових технологій зв'язку [12]. Він також визначається як мережа фізичних об'єктів або речей, наділених обмеженими обчислювальними та комунікаційними можливостями, а також вбудована в електроніку (наприклад, датчики та виконавчі механізми), програмне забезпечення та мережеве підключення, що дозволяє цим об'єктам збирати, інколи обробляти та обмінюватися даними. Речі IoT стосуються предметів повсякденного життя, починаючи від розумних побутових пристроїв, таких як розумна лампочка, інтелектуальний адаптер, розумний лічильник, розумний холодильник, розумна піч, змінний струм, датчик температури, детектор диму, IP-камера, до більш складних такі пристрої, як пристрої ідентифікації радіочастот (RFID), детектори серцебиття, акселерометри, датчики на стоянці та цілий ряд інших датчиків у автомобілях тощо [13]. IoT пропонує безліч додатків та послуг, починаючи від критичної інфраструктури, закінчуючи сільським господарством, військовою технікою, побутовою технікою та особистим медичним обслуговуванням [14]. Крім того, сфери, охоплені послугами IoT, включають, але не обмежуються цим, енергетику, управління будівлями, медицину, роздрібну торгівлю, транспорт, виробництво тощо. Величезний масштаб мереж IoT приносить нові завдання, такі як управління цими пристроями, величезний обсяг даних, зберігання, зв’язок, обчислення, безпека та конфіденційність. Наріжним каменем комерціалізації технологій IoT є гарантія безпеки та конфіденційності, а також задоволення споживачів. Той факт, що IoT використовує такі активні технології, як Software Defined Networking (SDN), хмарні обчислення та туманні обчислення, також збільшує масштаби загроз.

[2] https://www.toptal.com/designers/interactive/smart-home-domestic-internet-of-things

Дані, що генеруються пристроями IoT, є масовими, тому традиційні методи збору, зберігання та обробки даних можуть не працювати в такому масштабі. Крім того, величезна кількість даних може також використовуватися для моделей, поведінки, прогнозування та оцінки. Крім того, неоднорідність даних, сформованих IoT, створює ще один фронт для поточних механізмів обробки даних. Тому, щоб використати значення даних, сформованих IoT, потрібні нові механізми. У цьому контексті машинне навчання вважається однією з найбільш підходящих обчислювальних парадигм для забезпечення вбудованого інтелекту в пристроях IoT. Машинне навчання може допомогти машинам та інтелектуальним пристроям зробити висновок про знання з пристрою чи даних, створених людиною. Це також можна визначити як здатність інтелектуального пристрою змінювати або автоматизувати ситуацію чи поведінку на основі знань, які вважаються важливою частиною рішення IoT. Методи машинного навчання використовувались у таких завданнях, як класифікація, регресія та оцінка щільності. Різноманітні додатки, такі як комп'ютерний зір, виявлення вторгнень, біоінформатика, виявлення шкідливого програмного забезпечення, аутентифікація та розпізнавання мови, використовують алгоритми та методи машинного навчання. Подібним чином машинне навчання можна використовувати в IoT для надання інтелектуальних послуг. Однак дане дослідження зосереджено на застосуванні машинного навчання для надання послуг безпеки та конфіденційності мережам IoT.

Конкуренція на ринку та технічні обмеження є перешкодою для підвищення безпеки цих пристроїв. Що ще гірше, найчастіше імена користувачів та паролі за замовчуванням не змінюються, що робить ці пристрої головною мішенню для використання противниками. Це було показано в останні роки при широкомасштабній DDoS-атаці на інфраструктури масштабу Інтернету. Нові ботнети, також показують, як супротивники постійно адаптуються, щоб уникнути виявлення, і залишаються постійною загрозою для поступово зростаючого діапазону IoT. Ці бот-мережі IoT також сканують Інтернет на наявність інших пристроїв, використовуючи налаштування за замовчуванням від виробника, і можуть швидко перерости у потужну збірку зброї, що спричинить серйозні наслідки для багатьох зацікавлених сторін.

Зараз «дірки» в Інтернеті або «темні мережі» представляють невикористані адреси без законного трафіку, спрямованого на них. Аналіз цього фонового випромінювання в Інтернеті, що складається з переважно небажаного трафіку, може, таким чином, дати ключове уявлення про методологію зловмисників і стати кроком до розробки ефективних моделей, які можуть ідентифікувати та повідомляти про вхідне сканування чи інші шкідливі дії.

Загалом, першим кроком на шляху захисту є аналіз та класифікація характеристик мережі шкідливих пристроїв. Цей аналіз може допомогти у побудові моделей, які можуть бути розгорнуті на периметрах підприємства та Інтернет-провайдера для виявлення підозрілої діяльності в організації.

Традиційні моделі машинного навчання спрямовані на:

– характеристику та визначення рівня шкідливих дій пристроїв IoT, наприклад, сканування та атаки DoS, на основі характеристик мережі, отриманих від мережевих телескопів;

– навчання та тестування нейронної мережі для класифікації пристроїв IoT на основі особливостей мережевого трафіку для забезпечення класифікації IP-адрес, близьких до реального часу;

– оцінки ефективності алгоритмів.

Традиційні алгоритми машинного навчання можна класифікувати на три категорії: алгоритми навчання з вчителем, без вчителя та підкріплення.

Навчання з вчителем: Навчання з вчителем виконується, коли визначено конкретні цілі, які охоплюють певний набір вхідних даних. Для цього типу навчання дані спочатку маркуються, а потім проводиться навчання з маркованими даними (що мають входи та бажані результати). Воно намагається автоматично визначити правила з доступних наборів даних та визначити різні класи, і нарешті передбачити належність елементів (об'єктів, індивідів та критеріїв) до даного класу.

Навчання без вчителя: При навчанні без вчителя середовище забезпечує лише вхідні дані без бажаних цілей. Воно не вимагає маркованих даних і може досліджувати схожість між маркованими даними та класифікувати дані за різними групами.

Навчання під контролем та неконтрольовані прийоми в основному зосереджуються на проблемах аналізу даних, тоді як навчання з підкріплення переважно для порівняння та прийняття рішень. Ця категоризація та вибір методики машинного навчання визначаються залежно від характеру наявних даних. Коли тип вхідних даних та бажані результати (мітки) відомі, використовується навчання з вчителем тобто контрольоване навчання. У цій ситуації система

навчена лише прив'язувати входи до бажаних виходів. Класифікація та регресія є прикладами контрольованих технік навчання, коли регресія працює з безперервними, а класифікаційні роботи з дискретними результатами. Різноманітні методи регресії, такі як підтримка векторної регресії, лінійна регресія та поліноміальна регресія є загальновживаними методами. З іншого боку, класифікація працює з дискретними вихідними значеннями (мітки класів). Поширені приклади алгоритмів класифікації включають K-найближчого сусіда, логістичну регресію та метод опорних векторів. Деякі алгоритми можуть бути використані як для класифікації, так і для регресії, такі як нейронні мережі. Коли результати не є чітко визначеними, і система повинна виявити структуру в межах вихідних даних, для навчання системи використовуються методи контролю без вчителя. Навчання без вчителя включає кластеризацію, яка об'єднує об'єкти на основі встановлених критеріїв подібності, таких як кластеризація K-засобів. Ступінь точності передбачувальної аналітики залежить від того, наскільки відповідний метод машинного навчання використовував минулі дані для розробки моделей, і наскільки добре він спрогнозував майбутні значення. Для прогнозного моделювання використовуються такі алгоритми, як метод опорних векторів, нейронні мережі та Наївний баєсів класифікатор.

Навчання з підкріпленням: Навчання з підкріпленням не визначає конкретних результатів, і агент вчиться на основі зворотного зв'язку після взаємодії з навколишнім середовищем. Він виконує деякі дії та приймає рішення на основі отриманої винагороди. Агент може бути винагороджений за вчинення хороших дій або покарання за погані дії та використовувати критерії зворотного зв'язку, щоб максимізувати довгострокові винагороди.

Глибоке навчання – це техніка машинного навчання. Нейронна мережа складається з нейронів (розглядаються як змінні), з'єднаних через зважені зв'язки (розглядаються як параметри). Для досягнення бажаного набору результатів, під наглядом або без нагляду техніка навчання пов'язана з мережею.

Навчання здійснюється з використанням маркованих та немаркованих даних із контрольованих або неконтрольованих технік навчання, відповідно з подальшим ітеративним регулюванням ваг серед кожної пари нейронів.

Глибоке навчання відоме розподіленими обчисленнями, а також вивченням та аналізом величезної кількості немаркованих, некатегоризованих та неконтрольованих даних. Воно розробляє ієрархічну модель навчання та представлення особливостей, мотивованих багатошаровим процесом навчання в мозку людини. Моделі глибокого навчання сприяють різним програмам машинного навчання, таким як розпізнавання мови, комп'ютерний зір та обробка природної мови, забезпечуючи вдосконалене моделювання класифікації та генеруючи кращі зразки даних. Крім того, ці моделі також корисні для стиснення та відновлення даних як у часових, так і в просторових областях завдяки своїй ефективності у витягуванні шаблонів та функцій із великих обсягів даних та витягуванні відносин у залежних від часу даних.

Навчання глибокому підкріпленню: глибоке навчання – це один із методів наближення функцій, класифікація та прогнозування функцій, тоді як навчання з підкріпленням – це інший тип методів машинного навчання, що застосовуються для прийняття рішень, коли агент програмного забезпечення дізнається про оптимальні дії, взаємодіючи із середовищем у різних станах. Глибоке навчання і навчання з підкріпленням вступають у гру разом у ситуації, коли кількість станів та розмірність даних дуже великі, а середовище нестаціонарне. Тому традиційне навчання з підкріпленням недостатньо ефективне. Поєднуючи глибоке навчання і навчання з підкріпленням, агенти можуть вчитися самостійно та формувати дієву політику для отримання максимальних довгострокових винагород. У цьому підході навчання з підкріпленням отримує допомогу від глибокого навчання, щоб знайти найкращу політику, і глибоке навчання виконує апроксимацію значень дії, щоб знайти якість дії в даному стані.

Традиційні методи машинного навчання, що використовуються в безпеці IoT працюють з маркованими даними і використовуються в мережах IoT для зондування спектру, оцінки каналів, адаптивного фільтрування, проблем безпеки та локалізації. Ця категорія містить два різних типи методів: класифікація та регресія. Класифікація під контрольованим машинним навчанням використовується для прогнозування, а також моделювання наявних наборів даних. Регресія використовується для прогнозування безперервних числових змінних.

Сімейство неконтрольованих алгоритмів навчання має справу з неміченими даними та використовує вхідні дані евристично. Вони використовуються для виявлення аномалій, несправностей та вторгнень, кластеризації комірок та балансування навантаження. Кластеризація за категорією безконтрольного навчання використовується для групування даних на основі деяких

притаманних подібностей та відмінностей. Кластеризація не контролюється, а отже, немає правильних чи неправильних відповідей. Для оцінки точності результатів використовується візуалізація даних.

Відсутність емпіричних даних для атакуючих пристроїв IoT є ключовим недоліком багатьох дослідницьких заходів (через різні проблеми конфіденційності та логістики при спробі отримати таку інформацію з місцевих сфер IoT). Це обмеження актуальне у контексті досліджень машинного навчання, оскільки воно покладається на достатню репрезентацію аномальної поведінки для прийняття надійних рішень в реальних умовах.

Щоб вирішити цю проблему, необхідним є використання мережевих телескопів, щоб отримати уявлення про небезпеку пристроїв IoT в Інтернеті, тобто за допомогою набору даних Міжнародної асоціації з аналізу Інтернет-даних «CAIDA»[3]. На цьому об'єкті функціонує інфраструктура збору, курирування та розподілу даних та надає її науковому та дослідницькому співтовариству за запитом. Великий архів вимірювань CAIDA вирішує проблему відсутності достатньої кількості даних для аналізу алгоритмів машинного навчання. Основна причина вибору інформації заголовка IP для вилучення функцій пов’язана з тим, що отримання та обробка цієї інформації займає менше часу та ресурсів, ніж аналіз всього корисного навантаження. Крім того, пристрої IoT, що використовуються як частина бот-мережі, матимуть подібні характеристики мережі, коли вони намагаються сканувати Інтернет на наявність інших вразливих машин.

Загалом, методи машинного навчання спрямовані на аналіз даних та вилучення з них корисних висновків. Взагалі, набір векторів даних, що називаються функціями, служить вхідними даними для алгоритмів, і результати в значній мірі залежать від типу навчання, що виконується. Як результат, схеми машинного навчання можуть надати раніше невидиму інформацію про дані. Зараз вибір найбільш важливих функцій є, мабуть, одним з найважливіших кроків у будь-якому проекті машинного навчання. Наприклад, некоректні або зайві функції можуть суттєво вплинути на точність алгоритму машинного навчання. Отже, дослідження класифікації мережевого трафіку вже зосереджено на виявленні потоків із необроблених даних та вилученні функцій на основі цих потоків.

Зараз даркнет, як уже згадувалося раніше, не має фізичних пристроїв, пов’язаних із виділеними IP-адресами. Таким чином, будь-який спрямований на нього трафік не отримує відповіді. Тому поняття потоків, визначене для звичайної класифікації інтернет-трафіку, тут не відповідає дійсності. Отже, розглядаємо потік, який ідентифікується за IP-адресою джерела та витягує функції з декількох пакетів протягом заданої тривалості часу для кожного. По суті ці вектори функцій індексуються IP-адресою джерела. Загалом, статистичні особливості класифікації мережевого трафіку виявились дуже корисними для виявлення різних видимих і невидимих програм в Інтернеті, включаючи класифікацію пристроїв IoT.

Після вилучення даних позначаємо дані пристроїв IoT, знайдених у даркнеті. Дані даркнет є унікальними в тому сенсі, що це односторонній зв'язок (без можливої відповіді від нерозподіленої IP-адреси, знайденої там). Цей факт ускладнює використання певних традиційних підходів до маркування даних.

Небезпечні пристрої IoT постійно намагаються заразити інші пристрої в Інтернеті, щоб збільшити розмір їх ботнету і, отже, посилити вплив DoS-атак, організованих за допомогою цих ботів. Такі пристрої можуть виконувати різні типи сканування, включаючи, але не обмежуючись, мережеве сканування, сканування портів, тощо. Під час сканування портів атакуючий намагається ідентифікувати активні порти на хості, надсилаючи пакети запитів з метою використання вразливості, пов'язаної зі службою, що працює на цьому порту. Тому, якщо виявлено, що конкретна вихідна IP-адреса націлена на понад 5 портів для кожної унікальної IP-адреси призначення, а також надсилає понад 50% її пакетів, позначаємо це як сканування портів.

Під час сканування мережі зловмисник прагне використовувати одну службу, націлюючи один і той же порт на кілька хостів. Ці скани часто використовуються для розширення мережі ботів шляхом набору більшої кількості пристроїв.

Нарешті, стелс-скани часто використовуються зловмисниками, щоб уникнути систем запобігання / виявлення вторгнень. Ці сканування передбачають надсилання невеликої кількості TCP або UDP-пакетів за короткий проміжок часу невеликій кількості хостів. Розглядаємо вихідну IP-адресу як виконуючу стелс-сканування, якщо вона відправляє менше 15 пакетів менше ніж на 15 хостів, націлених на менш ніж 5 портів за певний часовий проміжок.

[3] https://www.caida.org/

Підробка IP-адрес – це загальний прийом, який застосовується супротивниками в контексті DDoS-атак. Таким чином, коли ціль заповнена запитами, вона надсилає відповідь назад на ці підроблені IP-адреси. Однак це не призведе до відповідей, якщо цільові IP-адреси перебувають у даркнеті, тобто відсутні фізичні пристрої, пов'язані з цими IP-адресами. Тому будь-який пакет відповідей, побачений у даркнеті, є чітким показником зворотного розсіювання трафіку від DDoS-атаки. Відповідно, позначаємо вихідну IP-адресу «жертвою» DDoS-атаки, якщо бачимо з неї пакет відповідей у мережі. Будь-який TCP-пакет, що має встановлений прапор RST, також ідентифікується як пакет відповіді. Нарешті, будь-які інші IP-пакети, побачені на пристроях IoT у даркнет, позначаються як неправильна конфігурація.

Класифікація трафіку допомагає операторам та постачальникам послуг у багатьох доменах, наприклад, якість обслуговування, моніторинг, виявлення вторгнень тощо. Однак пошук правильної гіпотези для прогнозування є суттю мети керованого алгоритму машинного навчання.

Трафік IoT зазвичай характеризується великим обсягом, різноманітністю, змінною швидкістю та невизначеністю. Більшість традиційних методів машинного навчання не є ефективними та масштабованими для управління даними IoT, тому потребують значних модифікацій. Більше того, невід'ємні невизначеності існують у даних IoT, і їх важко позбавити цієї внутрішньої непередбачуваності.

Загальні обмеження використання методів машинного навчання у мережах IoT.

1) Обробна потужність та енергія: Алгоритми машинного навчання за своєю суттю мають деякі складності, такі як пам'ять, обчислювальна робота та складність вибірки. Крім того, звичайні підходи машинного навчання не мають масштабованості і обмежуються лише проблемами з низькими розмірами. Пристрої IoT невеликі і зазвичай мають обмеження енергії з обмеженою обробною потужністю. Отже, безпосереднє застосування звичайних методів машинного навчання не підходить у середовищах, обмежених ресурсами. З іншого боку, розумні пристрої IoT вимагають обробки даних у реальному часі для додатків реального часу, тоді як традиційні методи машинного навчання не призначені для обробки постійних потоків даних у режимі реального часу. Після таких обмежень необхідно об'єднати існуючі потокові рішення з алгоритмами машинного навчання; однак це збільшить загальну складність алгоритму.

На додаток до цього, мережі, засновані на машинному навчанні, розробляються, припускаючи, що весь набір даних доступний для обробки на етапі навчання. Однак це не відповідає даним IoT. Це явище породжує різні проблеми, коли традиційні методи повинні обробляти безпрецедентний обсяг даних. Також прогнозована здатність алгоритму зменшується зі збільшенням розмірності даних. Попереднє обговорення, зокрема, стосується функцій, пов'язаних із безпекою в IoT, де дані в реальному часі обробляються для можливих векторів атак, таких як вторгнення тощо.

2) Управління даними та аналітика: бездротові дані можуть генеруватися з різних джерел, включаючи мережеві інформаційні системи, а також пристрої зондування та зв'язку. Дані є першопричиною для систем IoT, де необхідно проводити ефективний аналіз для отримання значущої інформації з даних; однак масове управління даними є серйозною проблемою в Інтернеті речей з усіх стандартних програм. Дані, що генеруються в мережах IoT, різноманітні за своїм характером, різними типами, форматами та семантикою, таким чином демонструючи синтаксичну та семантичну неоднорідність. Синтаксична неоднорідність відноситься до різноманітності типів даних, форматів файлів, схем кодування та моделей даних. Тоді як семантична неоднорідність відноситься до відмінностей у значеннях та інтерпретаціях даних. Така неоднорідність призводить до проблем з точки зору ефективного та уніфікованого узагальнення, особливо у випадку великих даних та різних наборів даних з різними атрибутами.

Машинне навчання передбачає, що статистичні властивості у цілому наборі даних залишаються незмінними, і вимагає попередньої обробки та очищення даних перед тим, як помістити їх у певну модель. Однак це не так у реальному світі, коли дані з різних джерел мають різне форматування та подання. Крім того, між різними частинами одного і того ж набору даних також можуть бути відмінності. Така ситуація створює труднощі для алгоритмів машинного навчання, оскільки алгоритми, як правило, не призначені для обробки семантично та синтаксично різноманітних даних. Це явище виступає за ефективне вирішення проблеми неоднорідності.

**Висновки і перспективи подальших досліджень.** У роботі представлено традиційні моделі, які використовують методи машинного навчання для класифікації зловмисного трафіку Інтернету речей як виконуючого сканування, DoS-атак або як неправильну конфігурацію. Зокрема, вектори функцій тут базуються на статистичних властивостях, отриманих із даних даркнету.

Однак одне з обмежень використання методів машинного навчання у мережах IoT полягає в тому, що традиційні моделі навчаються виключно на даних даркнету, які в основному представляють мережеві атаки. Тому, перспективи подальших досліджень ґрунтуються на розробці реалістичної установки, яка імітує типове середовище IoT, тобто з використанням як нешкідливих, так і шкідливих даних IoT для навчання та тестування алгоритмів машинного навчання.

**Література:**

1. Петренко А. І. Криптологія в Інтернеті речей / А. І. Петренко // Моделювання та інформаційні системи в економіці. – 2019. – № 97. – С. 155-163. – Режим доступу: http://nbuv.gov.ua/UJRN/Mise_2019_97_18.
2. Семеног А. Ю. Цифрові технології в умовах формування цифрової економіки. Наукові записки Національного університету «Острозька академія». Серія «Економіка» : науковий журнал. Острог : Вид-во НаУОА, вересень 2020. –№ 19(47). –С. 20–28.
3. Інтернет речей, «Великі дані» та аналітичне опрацювання в «Розумному місті» / О. Забігайло, О. Терешко, І. Панасюк, Р. Леськів // ІМСТ, 11-12 грудня 2019 року. – Т.: ТНТУ, 2019. – С. 45.
4. Проблеми та загрози IoT пристроїв / І. Р. Опірський, Р. В. Головчак, І. Р. Мойсійчук, Т. С. Балянда, С. П. Гаранюк // Кібербезпека: освіта, наука, техніка: Київський університет імені Бориса Грінченка, 2021. – № 3 (11). – С. 31-42.
5. Косяков Д.Г., Ларченко О.В. Тенденції розвитку сучасних інформаційних технологій. Сучасна молодь в світі Інформаційних технологій: матеріали ІІ Всеукраїнської науково-практична інтернет-конференції молодих вчених та здобувачів вищої освіти присвячена Дню науки. 14 травня 2021 р. Херсон: ХДАЕУ. – С.53-54.
6. Yeboah-Ofori, A., Islam, S., & Yeboah-Boateng, E. Cyber Threat Intelligence for Improving Cyber Supply Chain Security. In 2019 International Conference on Cyber Security and Internet of Things (ICSIoT), IEEE. – 2019. – pp. 28-33.
7. Smith, A. (2020, 2 лютого). The Five Biggest Security Threats and Challenges for IoT - DZone IoT. dzone.com. https://dzone.com/articles/the-biggest-security-threats-and-challenges-for-io
8. Razali, M. F., Razali, M. N., Mansor, F. Z., Muruti, G., & Jamil, N. (2018). IoT Honeypot: A Review Information and Network Security (AINS).doi:10.1109/ains.2018.8631494
9. Schroer A. (2020). AI and the bottom line: 15 examples of artificial intelligence in finance. BuiltIn. March 25. (10 October 2020).
10. Rishi R., Saluja R. EY Future of IoT // Ernst & Young Associates LLP, 2019 – 32 с. Режим доступу: https://www.ey.com/ Publication/vwLUAssets/EY_-_Future_of_IoT/$FILE/EY-future-of-lot.pdf
11. O'Halloran D., D'Souza F. (2020). Data is the new gold. This is how it can benefit everyone – while harming no one. World Economic Forum. July 29. (03 October 2020).
12. Абрамов В. О. Методичні аспекти викладання дисциплін напрямку "інтернет речей" / В. О. Абрамов, О. С. Литвин // Кібербезпека: освіта, наука, техніка. – 2018. – № 1. – С. 73-85. – Режим доступу: http://nbuv.gov.ua/UJRN/cest_2018_1_10.
13. Праворська, Н.І. Розробка моделі елементів інтернету речей для контролю параметрів навколишнього середовища [Текст] / Н. І. Праворська // Вісник Хмельницького національного університету. Технічні науки. – 2019. – №6. – С. 116-119.
14. Самойленко М. Ю. Принципи застосування технології інтернет речей у сучасному світі техніки / М. Ю. Самойленко // Вчені записки ТНУ імені В.І. Вернадського. Серія: технічні науки, 2020. – Том 31 (70) Ч. 1 № 6. – С. 142-148.