ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА

на тему:

«АЛГОРИТМ ВИЯВЛЕННЯ АНОМАЛІЙ В СИСТЕМАХ ІНТЕРНЕТУ РЕЧЕЙ»

на здобуття освітнього ступеня магістр
за спеціальності 126 Інформаційні системи та технології
(код,найменування спеціальності)
освітньо-професійної програми Інформаційні системи та технології
(назва)

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Віталій БОРИСЮК
(підпис) *(ім'я, ПРІЗВИЩЕ здобувача)*

Виконав:
здобувач вищої освіти
група ІСДМ-63 Віталій БОРИСЮК
(ім'я, ПРІЗВИЩЕ)

Керівник: Олег СЕНЬКОВ
к.т.н. *(ім'я, ПРІЗВИЩЕ)*

Рецензент: _____
науковий ступінь, вчене звання *(ім'я, ПРІЗВИЩЕ)*

Київ 2024

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Навчально-науковий інститут Інформаційних технологій

Кафедра Інформаційних систем та технологій
Ступінь вищої освіти магістр
Спеціальність 126 Інформаційні системи та технології
Освітньо-професійна прорама Інформаційні системи та технології

ЗАТВЕРДЖУЮ
Завідувач кафедрою ІСТ
_____ Каміла СТОРЧАК

"____" _____2024 року

З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Борисюку Віталію Михайловичу
(прізвище, ім'я, по батькові здобувача)

1.Тема кваліфікаційної роботи: Алгоритм виявлення аномалій в системах Інтернету речей
керівник кваліфікаційної роботи Олег СЕНЬКОВ, к.т.н.
(ім'я, ПРІЗВИЩЕ, науковий ступінь, вчене звання)
затверджені наказом Державного університету інформаційно-комунікаційних технологій від "15" жовтня 2024 р. № 320

2. Строк подання кваліфікаційної роботи «27» грудня 2024 р.

3. Вихідні дані кваліфікаційної роботи:
1. Технології Інтернет речей.
2. Архітектура IoT.
3. Методи виявлення аномалій і IoT.
4. Науково-технічна література.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити):
1. Дослідження тенденцій розвитку та поширення Інтернет речей.
2. Огляд методів виявлення аномалій в IoT.
3. Аналіз результатів впровадження моделі виявлення аномалій в IoT.

5.Перелік ілюстраційного матеріалу: *презентація*

6. Дата видачі завдання «15» жовтня 2024р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1.	Підбір технічної літератури	15.10-05.11.24	
2.	Дослідження тенденцій розвитку та поширення Інтернет речей	06.11-12.11.24	
3.	Дослідження методів виявлення аномалій в IoT	13.11-19.11.24	
4.	Результати впровадження моделі виявлення аномалій в IoT	20.11-03.12.24	
5.	Висновки по роботі	04.12-10.12.24	
6.	Розробка демонстраційних матеріалів, доповідь.	11.12-20.12.24	
7.	Оформлення бакалаврської роботи	21.12-27.12.24	

Здобувач вищої освіти _____ Віталій БОРИСЮК
(підпис) *(ім'я, ПРІЗВИЩЕ)*

Керівник кваліфікаційної роботи _____ Олег СЕНЬКОВ
(підпис) *(ім'я, ПРІЗВИЩЕ)*

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття ступня магістр: 73 стор., 33 рис., 7 табл., 28 джерел.

Мета роботи – дослідження методів виявлення аномалій у системах IoT із використанням алгоритмів машинного навчання.

Об'єкт дослідження – процес виявлення аномалій у системах IoT.

Предмет дослідження – аномалії в системах IoT.

Короткий зміст роботи. У першому розділі магістерської роботи виконано аналіз особливостей використання IoT у різних галузях. Проаналізовано функціонування з іншими технологіями. Розглянуто тенденції розвитку, які відображають не лише поточний стан ринку, але й дають уявлення про майбутнє зростання та інновації.

Виконано огляд досліджень у сфері систем виявлення вторгнень у контексті Інтернету та, зокрема, в контексті Інтернету речей. Проаналізовано різні методи, які використовуються для досягнення основних результатів

У третьому розділі описується методологія, спосіб, за допомогою якого модель виявлення аномалій в IoT використовує комбінацію двох алгоритмів. Представлені результати запуску моделі для виявлення аномалій в IoT.

КЛЮЧОВІ СЛОВА: ІНТЕРНЕТ РЕЧЕЙ, ВИЯВЛЕННЯ АНАОМАЛІЙ, МОДЕЛЬ, МЕТОТ, АЛГОРИМ, МАШИННЕ НАВЧАННЯ, КЛАСТЕРИЗАЦІЯ, РЕРЕВО РІШЕНЬ, МІЖМАШИННА ВЗАЄМОДІЯ.

ABSTRACT

The text part of the qualifying work for obtaining a bachelor's degree: 73 pp., xx fig., 7 tables, 28 sources.

The purpose of the work is to study methods for detecting anomalies in IoT systems using machine learning algorithms.

The object of the study is the process of detecting anomalies in IoT systems.

The subject of the study is anomalies in IoT systems.

Summary of the work. The first section of the master's thesis analyzes the features of the use of IoT in various industries. The functioning with other technologies is analyzed. Development trends are considered, which reflect not only the current state of the market, but also give an idea of future growth and innovation.

A review of research in the field of intrusion detection systems in the context of the Internet and, in particular, in the context of the Internet of Things is performed. Various methods used to achieve the main results are analyzed.

The third section describes the methodology, the way in which the anomaly detection model in IoT uses a combination of two algorithms. The results of running the proposed model for anomaly detection in IoT are presented.

KEYWORDS: INTERNET OF THINGS, ANOMALY DETECTION, MODEL, METHOD, ALGORITHM, MACHINE LEARNING, CLUSTERIZATION, SOLUTION POOL, INTER-MACHINE INTERACTION.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

БД – база даних;

ПЗ – Програмне забезпечення;

ШІ - Штучний інтелект;

ADP - Anomalies Detection Process;

CART - Classification and Regression Tree;

DT - Decision tree;

GSM - Groupe Spécial Mobile;

EM - Expectation Maximization;

FP - False positive;

FN - False negative;

LTE - Long-Term Evolution;

ML - Machine Learning;

M2M - Machine-to-Machine;

IDS - Intrusion detection system;

IoT - Internet of things;

IPS - Intrusion prevention system;

IP - Internet Protocol;

ITU - International Telecommunication Union;

IWC - Inverse weight clustering;

TP - True positive;

TN - True negative;

RFID - Radio frequency identification;

WCDMA - Wideband Code Division Multiple Access;

WSN - Wireless sensory network.

ВСТУП

Актуальність теми. Інтернет речей (IoT) є однією з найбільш швидкозростаючих технологій за останні роки. Це технологія, за допомогою якої мільярди розумних об'єктів або пристроїв, відомих як «Речі», можуть використовувати кілька типів датчиків для збору різних типів даних про себе та/або навколишнє середовище та обмінюватися ними з уповноваженими сторонами для кількох цілей, наприклад контролю моніторинг промислових об'єктів або покращення бізнес-послуг чи функцій.

Поширеність об'єктів IoT скрізь полегшує життя, але також надає зловмисникам безпрецедентну кількість цілей для атаки. Збитки, спричинені атаками на об'єкти IoT, можуть варіюватися від незначних до суттєвих, залежно від типу даних, якими оперує пристрій. Наприклад, компрометація IoT-системи, такої як камера відеоспостереження, що фіксує місця злочинів, може дозволити злочинцям уникнути покарання.

Кількість кібератак на IoT-системи стабільно зростає. Оскільки такі системи складаються з пристроїв IoT, мереж та програмного забезпечення, кожен компонент є вразливим до різних видів атак. Деякі з них мають фізичний характер, тоді як інші спрямовані на логічні аспекти, такі як дані, що збираються пристроями IoT або передаються мережею. Частина атак орієнтована на використання стандартних налаштувань маршрутизації, через що дані можуть потрапляти до неавторизованих осіб. Інші загрози можуть включати впровадження підроблених IoT-пристроїв, які передають шкідливу інформацію всім іншим елементам мережі. Таким чином захист цих пристроїв від зловмисників і несанкціонованого доступу та модифікації є дуже важливим питанням, що робить дану тему актуальною.

Мета роботи – дослідження методів виявлення аномалій в системах IoT із використанням алгоритмів машинного навчання.

Для досягнення мети, у магістерській роботі успішно виконано наступні завдання:

- Дослідження тенденцій розвитку та поширення Інтернет речей;

- Огляд методів виявлення аномалій в IoT;

- Аналіз результатів впровадження моделі виявлення аномалій в IoT.

Об'єкт дослідження – процес виявлення аномалій в системах IoT.

Предмет дослідження – аномалії в системах IoT.

Методи дослідження. Під час написання магістерської кваліфікаційної роботи були використані методи теоретичного дослідження, імітаційного моделювання, методи виявлення на основі сигнатур, та методи виявлення аномалій.

Наукова новизна одержаних результатів. У ході дослідження описано новий підхід для виявлення аномалій у системах IoT на основі комбінації алгоритмів машинного навчання, інверсна вагова кластеризація (IWC) та алгоритм дерева рішень C4.5. Результат застосування яких показує високу точність виявлення аномалій у даних IoT.

Практична значущість одержаних результатів. Запропонована модель забезпечує ефективне рішення для виявлення аномалій у даних IoT на рівні додатків.

Апробація результатів магістерської роботи. Основні положення і результати магістерської роботи доповідались на науково практичних конференціях, що проходили на базі Державного університету інформаційно-комунікаційних технологій.

РОЗІДІЛ 1 ДОСЛІДЖЕННЯ ТЕНДЕНЦІЙ РОЗВИТКУ ТА ПОШИРЕННЯ ІНТЕРНЕТ РЕЧЕЙ

1.1 Огляд перспективних рішень застосувань Інтернету речей

Кількість підключених пристроїв IoT продовжує зростати експоненціально, що пояснює швидке розширення IoT. На даний момент понад 15 мільярдів активних підключень пристроїв IoT доводять домінування Інтернету речей у світі. Деякі з найкращих додатків IoT в охороні здоров'я та управлінні ланцюгами поставок заклали міцну основу для впровадження IoT.

Глобальний ринок Інтернету речей може досягти загальної капіталізації приблизно в 1,1 трильйона доларів за умови значного зростання витрат. Які сектори використовують постійно зростаючий обсяг коштів у глобальних оцінках витрат на IoT. Деякі з відомих варіантів використання IoT включають віддалений моніторинг і контроль активів, автоматизацію процесів, відстеження місцезнаходження, оптимізацію продуктивності активів і управління автопарком [1].

У звітах про дослідження ринку зазначено, що дохід від глобального ринку Інтернету речей може становити майже 293,2 мільярда доларів у 2025 році. Кількість пристроїв, підключених до Інтернету речей, може перевищити позначку в 16 мільярдів у 2025 році. Тому важливо знати про застосування Інтернету речей , які стане в центрі уваги в 2026 році. Огляд популярних випадків використання IoT може допомогти зрозуміти різні фактори, які сприяють застосуванню IoT.

Інтернет речей або IoT надає ефективне технологічне рішення для забезпечення інтеграції між реальним світом і комп'ютерними системами. Основна риса програми IoT виходить за рамки підключення, оскільки вона також допомагає забезпечити доступ до даних, що полегшує зв'язок між підключеними пристроями та користувачами. Наприклад, є можливість використовувати дані IoT, щоб зрозуміти, як працює конкретний пристрій у мережі, або отримати докладні відомості про споживання його заряду батареї.

За допомогою IoT кожен пристрій, підключений до Інтернету, може самостійно збирати, контролювати та обмінюватися даними. Уявіть собі вдосконалення традиційних операцій за допомогою пристроїв, які можуть виявляти аномалії у виробничих процесах і повідомляти точні дані тим, хто приймає рішення. Ось деякі з найважливіших реальних застосувань Інтернету речей, на які варто звернути увагу у 2025 році.

Рис.1.1 Сфери застосування IoT

Розумне місто. Smart City є цікавим прикладом реальних додатків IoT , які можуть створити міста майбутнього. Це революційна концепція застосування IoT для покращення міської інфраструктури. Використання систем на основі IoT може покращити ефективний доступ до комунальних послуг, якість і вартість життя разом з використанням наявних ресурсів [2]. Розумні міста також представляють

життєздатні перспективи для досягнення правдоподібних покращень економічного зростання разом із стимулюванням екологічної стійкості.

Основна передумова концепції «розумного міста» полягає в удосконаленні міського планування та державного управління шляхом використання автоматизації послуг. Розумні міста можуть допомогти у покращенні різних аспектів, таких як послуги управління, розподіл енергії та інноваційне міське сільське господарство. Переваги розумних міст також можуть запровадити життєздатні покращення в утилізації відходів, транспорті та управлінні дорожнім рухом разом із доступом до комунальних послуг, таких як вода та охорона здоров'я [3].

Розумні будинки. Якщо міста можуть стати «розумними» за допомогою IoT, чому б вам не думати так само щодо будинків у розумних містах? Популярні додатки в IoT також зосереджені на створенні розумних будинків, які спрямовані на підвищення якості життя. Розумні будинки передбачають використання IoT для керування різними побутовими приладами, сигналізацією, потоком води та освітленням. Крім того, концепція «розумного дому» також зосереджена на підвищенні безпеки ваших будинків за допомогою комплексних інтелектуальних систем безпеки на основі Інтернету речей.

Головною перевагою розумних будинків є гнучкість управління всіма процесами в будинку за допомогою смартфонів, ноутбуків і планшетів. Ви забули вимкнути ліфт перед виходом на роботу? Програми IoT можуть допомогти вам вимкнути світло у вашому домі за допомогою програми на вашому смартфоні. Найважливіше те, що можливість своєчасного сповіщення про роботу різних приладів у вашому домі може допомогти уникнути дорогого ремонту. У той же час розумні системи безпеки можуть захистити ваш будинок від різних загроз.

Розумні автономні транспортні засоби. Концепція безпілотних автомобілів здавалася темою з науково-фантастичних романів чи фільмів до появи IoT. Розумні автономні транспортні засоби – це найкраща відповідь на запитання « Яке майбутнє застосування технології IoT?» », оскільки вони можуть революціонізувати досвід користувачів транспортних засобів. На даний момент підключені пристрої в

автомобілях можуть допомогти контролювати певні функції. Наприклад, технологія GPS в автомобілях може допомогти в навігації до бажаних місць, а інтелектуальні пристрої можуть допомогти визначити оптимальний маршрут до місця призначення.

Використання IoT в розумних транспортних засобах також включає встановлення датчиків у автомобілі для контролю температури води в радіаторі або рівня моторного масла. Власники транспортних засобів також можуть стежити за загальним станом автомобіля, включаючи рівень палива, місцезнаходження та інші деталі, за допомогою програми на своїх смартфонах або ноутбуках.

Фітнес Wearables. Наступним помітним пунктом серед популярних додатків Інтернету речей у 2025 році будуть фітнес-носії. Фітнес-трекери на основі IoT можуть допомогти оптимізувати фітнес-цілі разом із відстеженням вашого прогресу в досягненні цілей. Список найпопулярніших додатків IoT для відстеження фізичної активності включає широкий спектр носіїв, які можуть допомогти в моніторингу щоденної діяльності. Ви можете вимірювати що завгодно – від режиму сну до рівня глюкози в організмі за допомогою одягу для фітнесу.

Найпоширеніший і популярний спосіб застосування Інтернету речей у відстеженні фізичної активності передбачає вимірювання статистики тренувань і кількості спалених калорій. Застосування IoT для відстеження фізичної форми також підкреслюють гнучкість використання переносних пристроїв і моніторингу вашого здоров'я. Наприклад, вам не потрібно приєднувати кілька проводів до тіла, щоб регулярно вимірювати пульс. Інтернет речей може допомогти вам вимірювати пульс протягом дня за допомогою розумного годинника [4].

Туризм і готельний бізнес на основі Інтернету речей. Різноманітність використання IoT у різних галузях пропонує багатообіцяючий простір для можливостей використання IoT у сфері гостинності. Інтернет речей представляє значний потенціал для оптимізації діяльності в індустрії гостинності та туризму. Найпоширеніша програма IoT у сфері гостинності зосереджена на оптимізації розподілу персоналу. Керівництво персоналу несе величезний тягар витрат на діяльність готелів, мотелів і курортів у сфері туризму та гостинності. IoT може

полегшити автоматизацію конкретних завдань, що може зменшити навантаження на персонал готелю.

Мобільні електронні ключі є одним із прикладів використання Інтернету речей у сфері гостинності та туризму. Ключі можуть допомогти гостям готелю отримати доступ до своїх номерів у готелі без взаємодії з будь-яким співробітником. Крім того, програми IoT також можуть допомогти гостям замовити обслуговування номерів або поділитися відгуками безпосередньо зі своїх смартфонів. У результаті керівництво готелю могло отримати корисну інформацію про поведінку та вимоги гостей за допомогою IoT.

Розумні мережі. Інтернет речей допомагає впроваджувати «розумні» в звичайні системи. Розумні мережі використовують можливості IoT для інноваційної трансформації інфраструктури розподілу енергії. Енергетичні компанії використовують додатки в IoT для створення розумних мереж, які можуть забезпечити кращу енергоефективність.

Інтелектуальні електромережі можуть надавати повні функціональні можливості для моніторингу споживання енергії разом із створенням прогнозів щодо відключень та дефіциту електроенергії. Крім того, інтелектуальні мережі також допомагають збирати дані про використання енергії різними компаніями та окремими особами. На індивідуальному рівні кожен дім може використовувати дані розумної електромережі для оцінки споживання енергії та визначення кращих способів енергозбереження [5].

Віддалений моніторинг пацієнтів. Найкориснішим реальним додатком Інтернету речей є віддалений моніторинг пацієнтів або RPM-системи з підтримкою IoT. Питання на кшталт « Яке майбутнє застосування технології IoT?» » привертають увагу до способів, якими IoT створює ціннісні переваги, які витримають випробування часом.

Ефективність технології Інтернету речей у вдосконаленні систем дистанційного моніторингу пацієнтів запровадила масштабні покращення у сфері охорони здоров’я. По суті, глобальна пандемія навчила важливості покращеної

модернізації та гнучкої доступності медичних послуг. Як пацієнти можуть отримати медичну допомогу, не відвідуючи своїх лікарів?

Інтернет речей дає життєздатну відповідь із використанням датчиків для моніторингу стану здоров'я пацієнтів. Датчики IoT можуть допомогти лікарям контролювати та відстежувати прогрес пацієнтів відповідно до рекомендованих планів лікування. Лікарі можуть використовувати просту програму IoT на своїх смартфонах, щоб перевіряти життєво важливий стан пацієнтів. Крім того, дистанційний моніторинг пацієнтів також може допомогти дізнатися, чи вчасно пацієнти приймають ліки. Застосування дистанційного моніторингу пацієнтів також може гарантувати, що пацієнти отримають своєчасну медичну допомогу в екстрених випадках.

Роздрібні програми IoT. Сектор роздрібної торгівлі також використовує IoT для досягнення кращої ефективності разом із стимулюванням нових підходів до інновацій. Роздрібний Інтернет речей може допомогти в забезпеченні покращення взаємодії з клієнтами разом із точністю відстеження продуктів. Найкращі додатки IoT у секторі роздрібної торгівлі також можуть підтримувати вдосконалення стратегії персоналу разом із керуванням запасами. Крім того, роздрібні торговці також можуть використовувати IoT для моніторингу моделей покупок клієнтів.

Дані IoT можуть допомогти роздрібним торговцям збирати інформацію про історію покупок, інформацію про місцезнаходження та тенденції купівельної поведінки. У результаті роздрібні торговці можуть у реальному часі отримувати інформацію про клієнтів, створюючи тим самим кращі стратегії для досягнення своїх бізнес-цілей.

Розумне виробництво.Промисловий IoT став однією з грізних галузей додатків Інтернету речей. Найбільш помітний вплив промислового Інтернету речей було виявлено у виробничому секторі. Одне з найбільш значущих застосувань в IoT для виробництва вказує на приклади розумних фабрик, які використовують технологію IoT.

Розумні фабрики можуть використовувати технологію Інтернету речей для збору даних про різні промислові пристрої та процеси. Спираючись на знання

промислових процесів і пристроїв, виробники можуть розробляти стратегічні плани, а також покращувати ефективність.

Застосування Інтернету речей у виробництві передбачає інтеграцію датчиків IoT з виробничим обладнанням та іншими фабричними активами. Найбільшою перевагою інтелектуального виробництва є профілактичне обслуговування, яке допомагає уникнути дорогого ремонту. Крім того, додатки IoT в інтелектуальному виробництві також можуть підтримувати ефективне використання аналітики для оптимізації споживання енергії. У той же час IoT також може запропонувати функції відстеження активів, які можуть зіграти вирішальну роль у підвищенні продуктивності процесів ланцюга постачання [3,4].

IoT Farming. Останнє доповнення до списку додатків Інтернету речей стосуватиметься багатообіцяючих випадків використання IoT у сільському господарстві. Інтелектуальні фермерські додатки IoT можуть допомогти оптимізувати різні операції на фермі, які потребують часу та зусиль. Інтелектуальні програми Інтернету речей створили революційний прогрес у сфері сільського господарства. Наприклад, програма IoT може допомогти фермерам визначити найкращий час для збирання врожаю. Датчики IoT можуть допомогти фермерам передбачити зміни погоди разом із виявленням коливань вологості ґрунту та концентрації поживних речовин.

Розумне фермерство через IoT також передбачає використання датчиків для моніторингу безпеки ферм. Крім того, розумні мітки для худоби можуть допомогти фермерам з перевагами моніторингу в реальному часі інформації про здоров’я тварин. Інші потенційні приклади використання IoT у сільському господарстві вказують на автономне сільськогосподарське обладнання.

1.2 Сучасний стан та тенденції розвитку IoT

Інтернет речей залишається головним пріоритетом. Згідно з останньою 148-сторінковою доповіддю «Стан IoT – весна 2024 року» , IoT залишається в трійці найбільших пріоритетів корпоративних технологій, тоді як штучний інтелект став

головним технологічним пріоритетом. У нещодавніх опитуваннях PWC , KPMG і BCG респонденти поставили IoT на друге або третє місце після штучного інтелекту з точки зору пріоритетності інвестицій у нові технології, причому штучний інтелект займав перше місце за всіма напрямками.

Останнє дослідження показує, що зростання ШІ є сильним напрямком для ринку Інтернету речей вартістю 236 мільярдів доларів, оскільки компанії все більше цікавляться як ШІ, так і Інтернетом речей у своїх організаціях. Одним із свідчення є аналіз IoT Analytics дзвінків про прибутки компаній: з третього кварталу 2022 року кількість згадок про ці дві технології в одному дзвінку про прибутки зросла на 61%.

Аналіз показує, що штучний інтелект — не єдина тенденція, яка сприятиме розвитку ринку IoT. На рис.1.2 представлені перспективні тенденцій, які розглянуті у даному підрозділі [6].

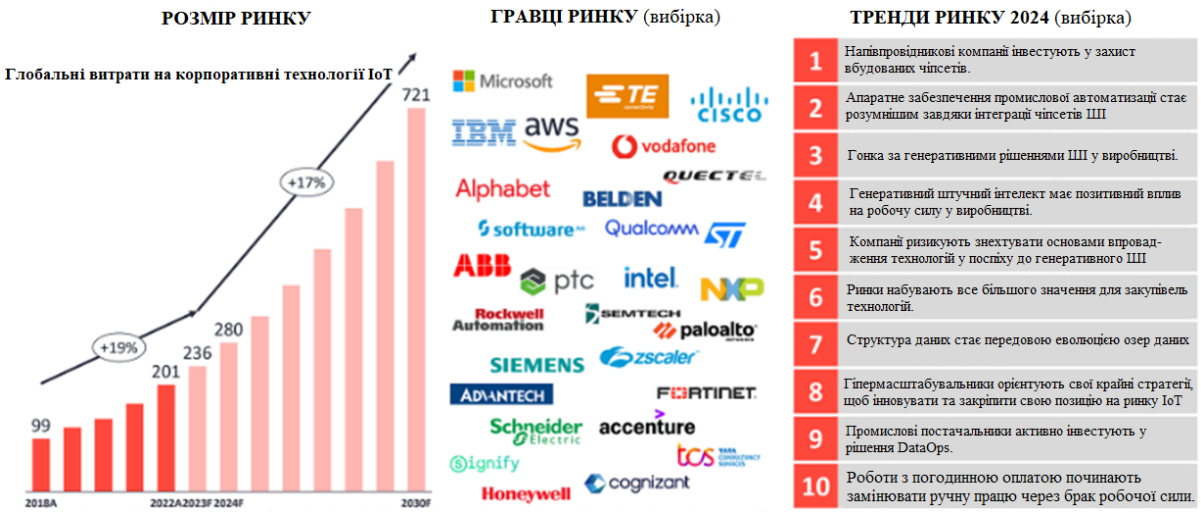

Рис. 1.2 Ринок IoT

1. Напівпровідникові компанії інвестують у захист вбудованих чіпсетів.

Напівпровідникові компанії все більше інвестують у захист вбудованих чіпсетів , щоб усунути зростаючі загрози безпеці, з якими стикаються пристрої IoT. Захист апаратного забезпечення на рівні чіпсета за допомогою захищених елементів і фізичних неклонованих функцій (PUF) може допомогти захистити дані, що надходять із периферійних пристроїв у хмару.

2. Апаратне забезпечення промислової автоматизації стає розумнішим завдяки інтеграції чіпсетів ШІ

З Розвитком технології, штучного інтелекту компанії тепер прагнуть використовувати штучний інтелект на периферії, збільшуючи попит на аналіз даних у реальному часі також на межі. Чіпсети штучного інтелекту також стають меншими за розміром, а потужність зростає. Ця тенденція призвела до появи IPC і шлюзів із вбудованими чіпсетами штучного інтелекту, що призвело до появи периферійного обладнання штучного інтелекту, яке може виконувати паралельні обчислення та навчати алгоритми з дуже низькою затримкою обчислювальної відповіді.

Однією з переваг використання чіпсетів AI на краю є прискорення обробки даних безпосередньо на промисловому обладнанні. Це, у свою чергу, зменшує мережевий трафік і підвищує безпеку, оскільки зменшується кількість даних, які надходять у хмару для обробки.

3. Гонка за генеративними рішеннями ШІ у виробництві. З'являється багато демонстрацій рішень на основі промислового генеративного штучного інтелекту (GenAI). Постачальники в промисловому та виробничому просторі поспішають розробити рішення на основі GenAI щодо кодування, усунення несправностей/підтримки, операційної аналітики та генеративного проектування, серед іншого.

4. Генеративний штучний інтелект має позитивний вплив на робочу силу у виробництві. Впровадження нових технологій у виробництво часто пов'язане з негативним впливом на робочу силу. Однак очікується, що впровадження GenAI у виробництво сприятиме підвищенню зайнятості та підвищенню кваліфікації, перемістивши фокус з автоматизації на стратегічне зростання. Згідно з McKinsey, оскільки GenAI щорічно вносить у світову економіку 2,6–4,4 трильйона доларів США , виробники, ймовірно, збільшать свої інвестиції в технології ШІ.

5. Компанії ризикують знехтувати основами впровадження технологій у поспіху до генеративного ШІ. Постачальники шукають способи застосувати це у своїх продуктах або створити нові, а кінцеві користувачі охоче приймуть це. Однак

цей поспіх не завжди корисний, коли йдеться про впровадження нових технологій. Ажіотаж часто може зрушити мислення як тих, хто прийняв рішення, так і постачальників.

6. Ринки набувають все більшого значення для закупівель технологій. Як компанії, так і продавці переходять на економіку передплати та прагнуть спростити процес закупівель. У січні 2024 року IoT Analytics опублікувала статтю, присвячену зростанню ринків B2B, зазначивши, що ринки B2B є каналом закупівель програмного забезпечення, який найшвидше розвивається.

7. Структура даних стає передовою еволюцією озер даних. Незважаючи на те, що це відносно новий термін, структура даних описує комплексну структуру інтеграції та управління даними. Він охоплює архітектуру, інструменти керування та спільні набори даних і призначений для допомоги організаціям у обробці їхніх даних.

8. Гіпермасштабувальники орієнтують свої крайні стратегії, щоб інновувати та закріпити свою позицію на ринку Інтернету речей. Хмарні провайдери стратегічно адаптуються до розвитку ринку IIoT. У нещодавніх розробках у сфері Інтернету речей відбулися важливі зміни серед основних провайдерів хмарних послуг, приділяючи більше уваги стратегіям периферії та контейнеризації.

9. Промислові постачальники активно інвестують у рішення DataOps. Кілька постачальників інвестують у промислові рішення DataOps для вирішення проблем інтеграції та аналізу даних. Industrial DataOps — це підхід до інтеграції даних, зосереджений на підвищенні якості даних за допомогою контекстуалізації та моделювання. Цей підхід привертає все більше уваги в промисловому просторі підключення.

10. Роботи з погодинною оплатою починають замінювати ручну працю через брак робочої сили. Компанії-виробники можуть отримати вигоду від обладнання як послуги (EaaS), замінивши операційні витрати, пов’язані з оплатою праці, іншим типом операційних витрат: робототехнікою як послугою (RaaS). RaaS — це відносно нова бізнес-модель, у якій робот надається машинобудівником на основі

результату (оплата за деталі, виготовлені з обладнанням) або на основі тривалості роботи (оплата за години використання обладнання), а не як пряма покупка.

1.3 Статистика, розміри ринку, використання та прогнози подальшого розвитку

У 2024 році кількість "розумних" будинків у США сягнула 69,91 мільйона. На 2025 рік цей показник зросте до 77,05 мільйона, а у 2026 році досягне 84,92 мільйона. До 2027 року 93,59 мільйона домогосподарств у США використовуватимуть розумні пристрої.

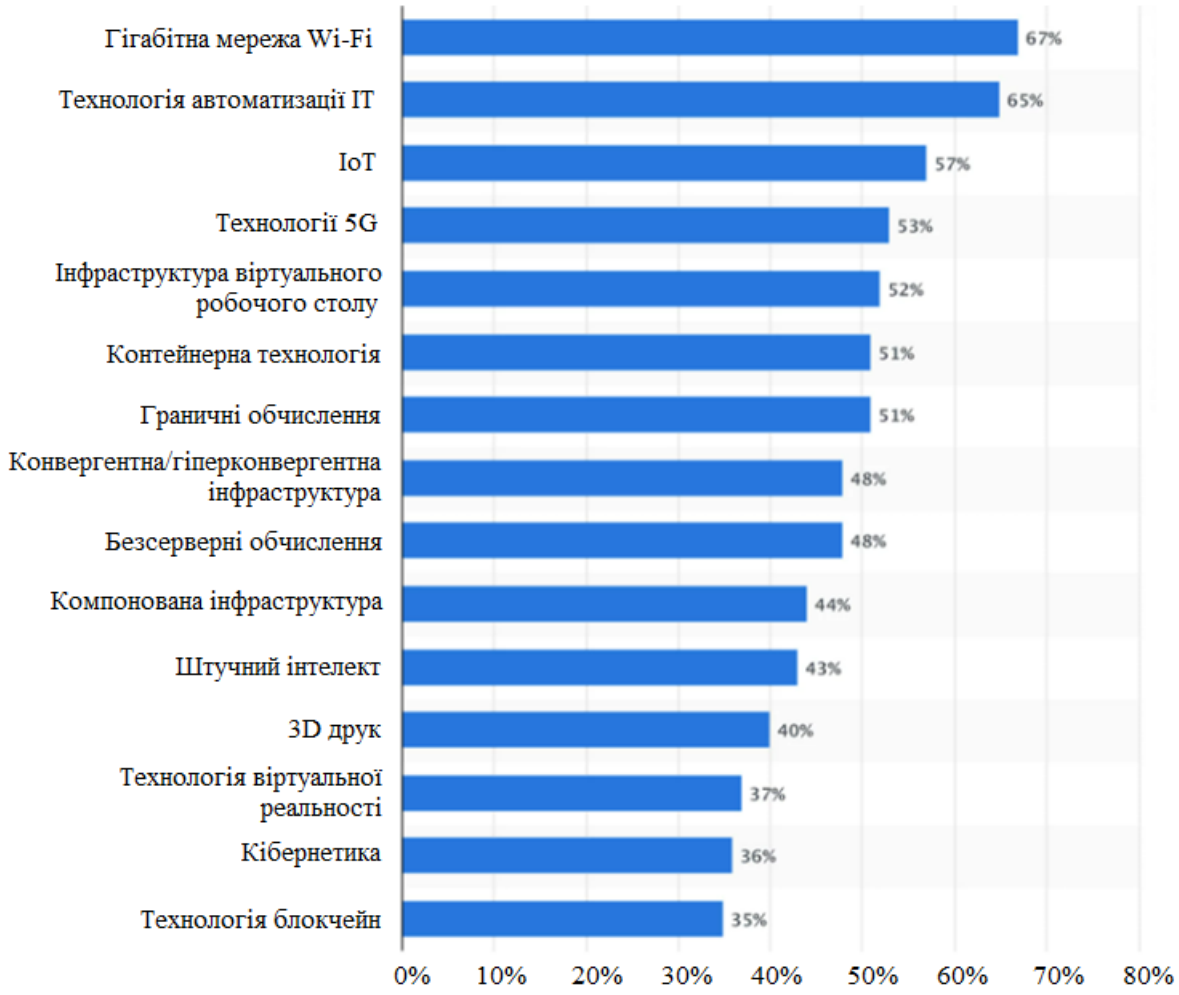

Рис.1.3 Розмір приросту пристроїв IoT

Очікується, що кількість пристроїв Інтернету речей (IoT) у світі майже подвоїться: з 15,1 мільярда у 2020 році до понад 29 мільярдів у 2030 році. — Statista

У 2023 році Інтернет речей став третьою за популярністю технологією серед впроваджених або запланованих до впровадження рішень у північноамериканських і європейських організаціях. Наразі 57% організацій з цих регіонів вже використовують IoT у своїй діяльності [7].

У 2021 році приблизно 41,9% домогосподарств використовували пристрої розумного дому, і прогнозується, що до 2025 року цей показник зросте до 48,4%.

У 2017 році лише 20% світового населення мав доступ до мобільного зв'язку, але в найближчі роки ця ситуація має значно змінитися.

У 2022 році розмір світового промислового ринку Інтернету речей становив 320,9 мільярда доларів США, і, згідно з прогнозами, до 2032 року він може досягти приблизно 1 562,35 мільярда доларів США, з середньорічним темпом зростання (CAGR) на рівні 17,2% у період з 2023 по 2032 рік [8].

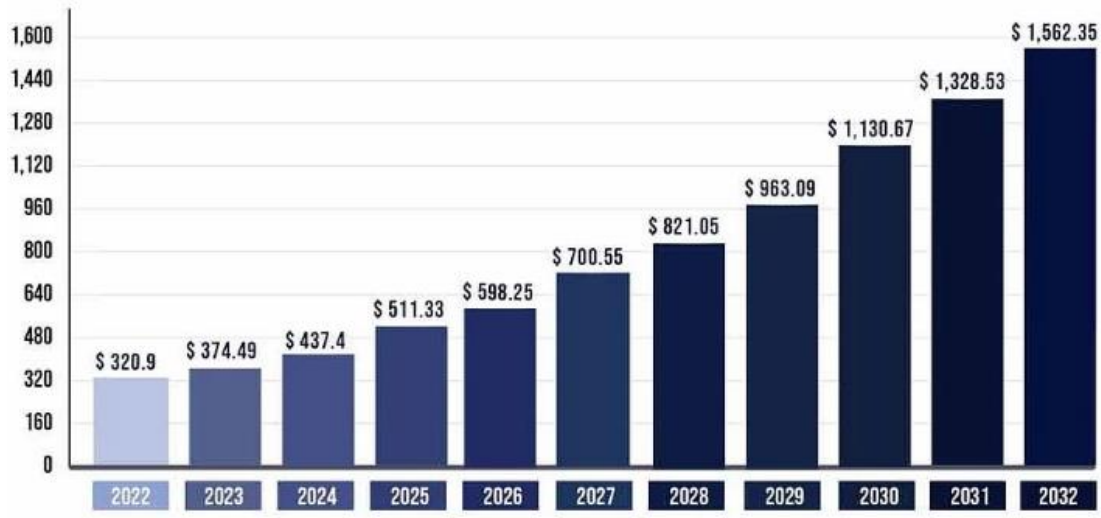

Рис.1.4 Розмір промислового ринку iot, 2022–2032 рр

До 2023 року США, Китай, Японія, Південна Корея та Німеччина очолять список топ-5 країн за кількістю пристроїв IoT та обсягом ринків IoT.

Інтернет речей може створити економічну вартість від 4 до 11 трильйонів доларів до 2025 року.

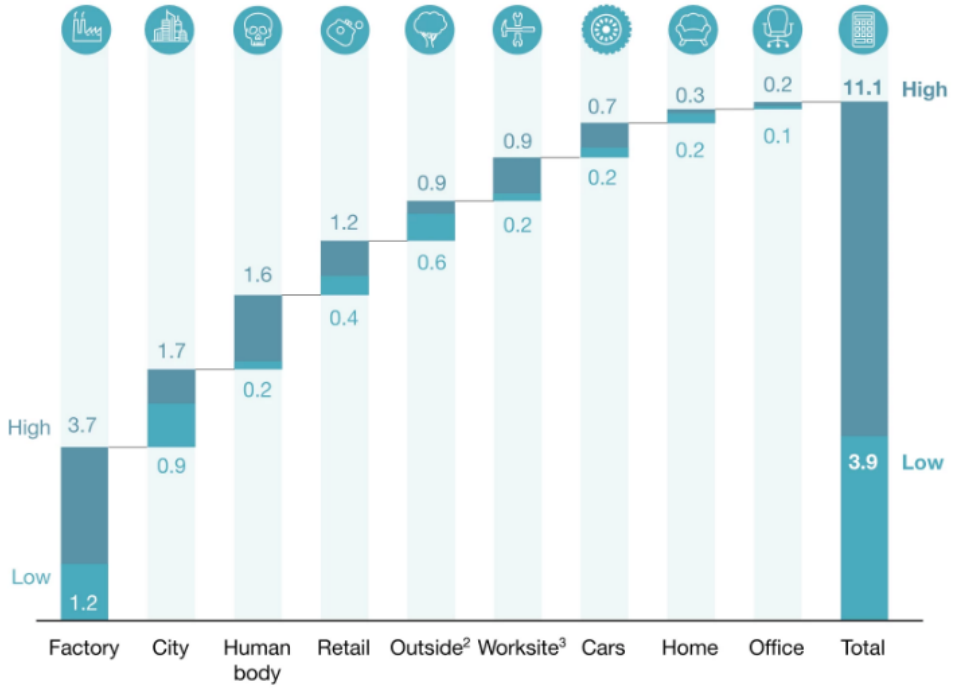

Рис.1.5 Економічна вартість

Безпека залишається одним із основних пріоритетів для розробників та користувачів IoT. Це пояснюється тим, що в 2021 році на пристрої Інтернету речей було здійснено 1,5 мільярда кібератак.

Згідно з даними IoT Analytics, найбільшу частину доходу від IoT-проектів приносить економія витрат, що є ключовим фактором для понад 50% бізнесів у цій сфері. Лише 35% таких проектів спрямовані на збільшення доходів.

58% виробників вважають, що Інтернет речей є важливою складовою для цифрової трансформації промислових процесів.

У сфері охорони здоров'я основною перевагою Інтернету речей є стимулювання інновацій. Наприклад, лічильники енергії, пристрої для отримання зображень та рентгенівські апарати використовують IoT. Найпоширенішим застосуванням є моніторинг стану пацієнтів.

У 2022 році виробничий сектор став лідером на ринку промислового Інтернету речей, завдяки широкому використанню IoT-рішень і цифрових технологій виробництва на підприємствах.

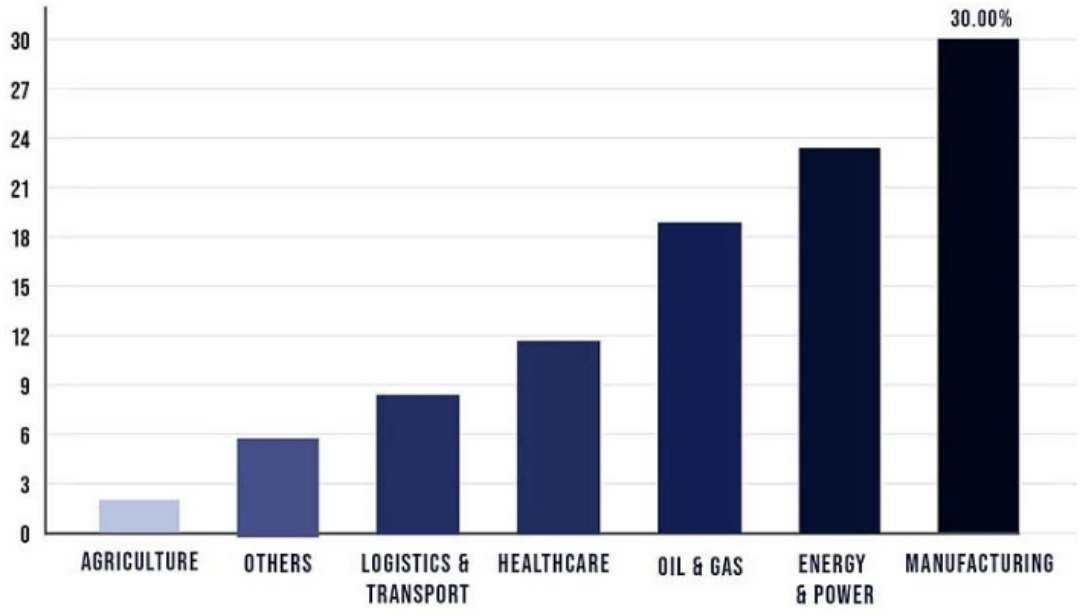

Рис.1.6 Лідируючі сектори на ринку промислового Інтернету речей

Щохвилини в Інтернеті з'являється 7 620 нових пристроїв Інтернету речей, тобто кожні 127 секунд з'являється новий пристрій. Прогнозується, що до 2025 року кількість підключених пристроїв на хвилину досягне 152 000.

Зараз світовий ринок Інтернету речей оцінюється в 662,21 мільярда доларів США, і, за прогнозами, до 2030 року він зросте до 3 352,97 мільярда доларів з середньорічним темпом зростання 26,1%.

Ринок сільськогосподарського Інтернету речей досяг 5,18 мільярда доларів до кінця 2023 року, а до 2025 року цей показник зросте до 6,98 мільярда доларів.

Зростання попиту на розумні пристрої в автомобілях сприяє розширенню світового ринку автомобільного Інтернету речей. Очікується, що до кінця 2025 року дохід від цього ринку досягне 397,2 мільярда доларів, а до 2028 року збільшиться до 882 мільярдів доларів.

У 2020 році витрати на IoT сягнули 749 мільярдів доларів, а в 2024 році вони, досягли 1 трильйона доларів. У 2025 році витрати, ймовірно, зростуть до 1,1 трильйона доларів і продовжуватимуть збільшуватись з кожним роком [9].

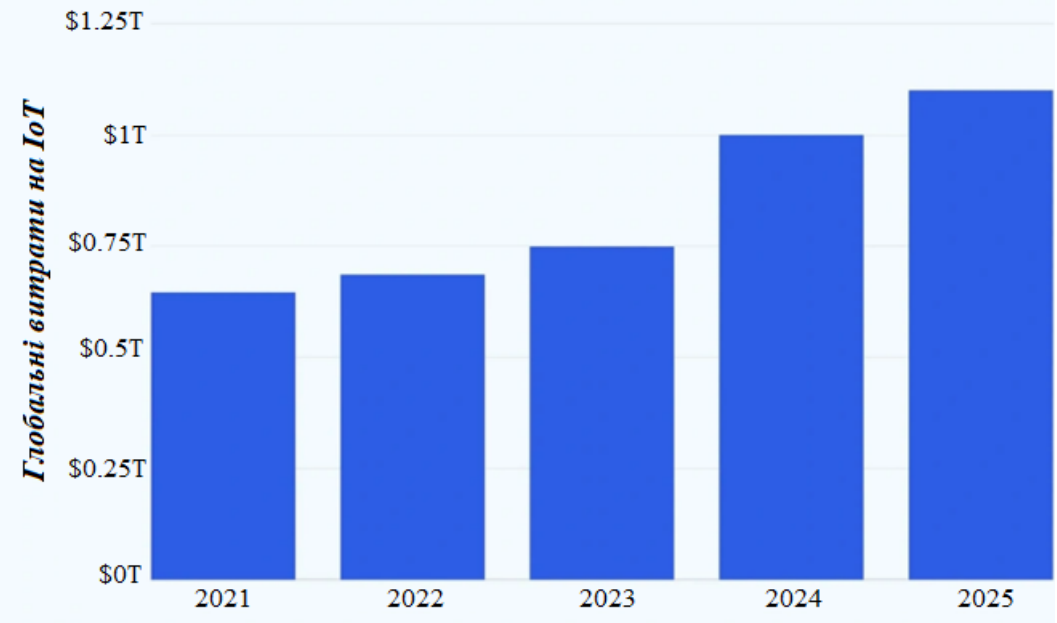

Рис.1.7 Щорічні витрати на IoT

У 2025 році пристрої Інтернету речей згенерують 79,4 зеттабайта (ZB) даних.

58% виробників IoT підтверджують, що використання IIoT у виробничих процесах є ключовим для успішної цифрової трансформації бізнесу.

США є лідером у впровадженні масштабних проектів з Інтернетом речей: понад 44% компаній повністю використовують цю технологію. За ними йдуть Великобританія, де 41% компаній інтегрують IoT, та Німеччина, де 35% компаній застосовують IoT.

1.4 Класифікація «речей» та архітектура IoT

Визначення IoT в IRC відноситься до «фізичних і віртуальних речей», які мають ідентичність, атрибути та характеристики. Аналіз показав, що деякими дослідниками IoT вказується як мережа для адресації фізичних об’єктів з точки зору зв’язку, контролю та повсюдності. Слід зазначити, що відображення між «речами» у фізичному та кіберсвіті є ядром IoT, і тому «речі» можна класифікувати на два типи: 1. Фізичні речі; 2. Кібер речі [10].

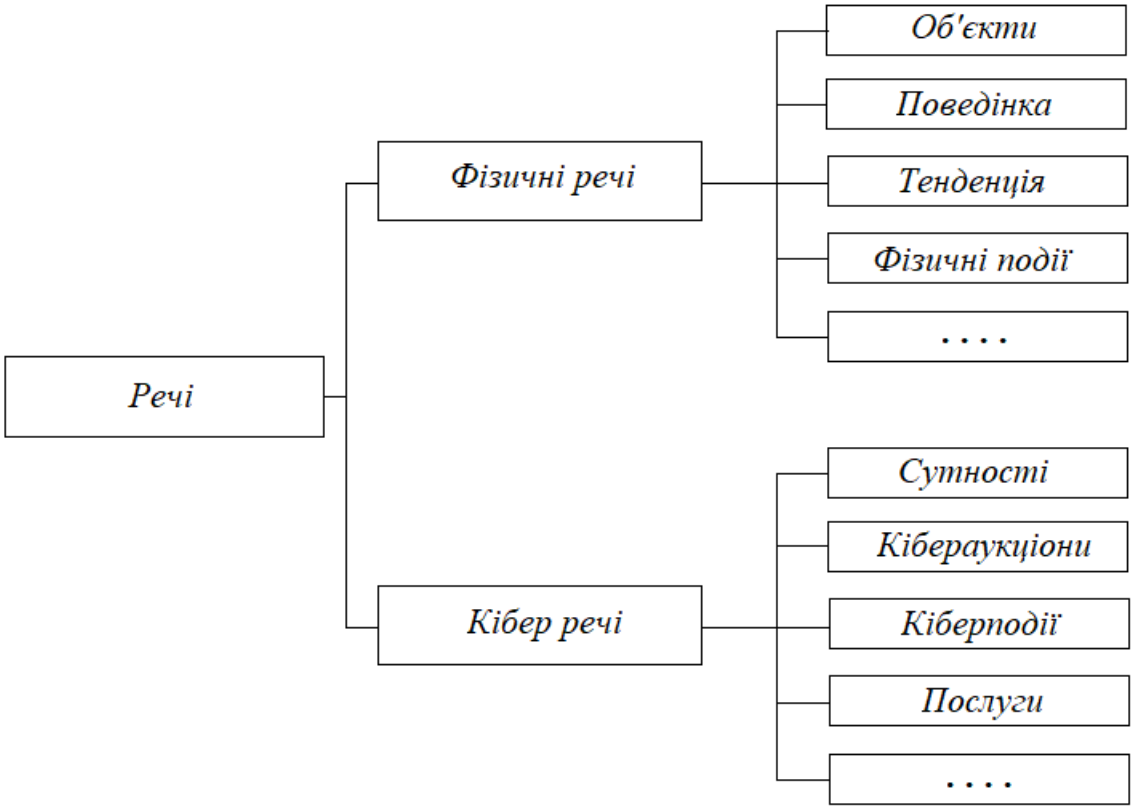

Рис.1.8 Класифікація речей

На рис.1.8 показано обидва типи «речей» в IoT. Фізичні речі охоплюють:

1. Об'єкти: які є конкретними речами з вимірними тілами. Прикладами об’єктів можуть бути транспортні засоби, люди, планшети та птахи.

2. Поведінка: це рухи об'єктів через певні мотивації. Водіння, біг, спостереження та їжа можуть бути прикладами поведінки об’єктів.

3. Тенденція: які є тенденції в самих речах або в результаті зовнішнього середовища, наприклад, затори на дорогах або дощова погода.

4. Фізичні події: це сукупність усіх вищезазначених, які разом описують те, що відбувається, спричинене певними обставинами у фізичному світі.

Кібер речі охоплюють:

1. Сутності: які вказують на абстрактні речі, такі як код і дані.

2. Дії: які вказують на обробку даних, наприклад передачу даних від однієї сутності до іншої.

3. Події: які вказують на причинно-наслідковий зв’язок дій, вжитих суб’єктом, наприклад, суб’єкт звітує про передачу даних.

4. Послуги: які вказують на функції, які пропонує річ, або функції, які пропонуються речі для досягнення конкретної мети. Фізичні та кібернетичні речі можна вважати машиною Google. Цей автомобіль представляє об’єкт, який має такі поведінки, як водіння, самообслуговування та паркування. Автомобіль може мати тенденцію, оскільки коли він виявляє, що погода почала йти дощем, очікується, що автомобіль виконає фізичні дії, такі як активація щіток склоочисника. Кіберречі привносять семантику та контекст у фізичні речі. Код, який розгортається на щітках склоочисників, представляє абстракції, які обмінюються та обробляють дані на основі послуг, доступних для виконання дій і звітування про події.

Архітектура IoT. Відповідно до досліджень, ITU запропонував багаторівневу архітектуру IoT. У цій архітектурі є чотири рівні; рівень пристроїв, мережевий рівень, рівень підтримки послуг і додатків і рівень додатків – рівні впорядковані за підходом «знизу вгору». На додаток до цих рівнів ITU запропонував два додаткові рівні; можливості управління та можливості безпеки. На рис.1.9 показано всі рівні архітектури, запропонованої ITU для IoT. Основні чотири шари, складені горизонтально, є функціональними шарами. Іншими словами, ці рівні використовуються для збору, передачі та обробки даних у мережах IoT. Інші два рівні, зображені вертикально, беруть частину функціональних можливостей кожного рівня з метою управління та безпеки. У нижній частині стека знаходиться «рівень пристроїв», який також відомий як «рівень сприйняття». На цьому рівні є пристрої IoT, такі як розумні електролічильники та датчики. Цей рівень також містить шлюзи, які є дуже важливим компонентом систем IoT. Шлюз IoT, як випливає з назви, — це пристрої, які отримують і збирають дані датчиків у межах свого діапазону прийому, транслюють протоколи датчиків, обробляють дані датчиків і направляють дані датчиків усередину або назовні [11].

У середовищі, повному датчиків, шлюзи відіграють життєво важливу роль. Уявіть собі будівлю із сотнями сенсорних об’єктів, які мають різні функції та можливості та виготовлені різними виробниками. Очікується, що ці об’єкти відчуватимуть широкий спектр тенденцій – напр. світло, температура, вологість,

шум, люди та вогонь, а також обмінюватися цими даними зі шлюзом за допомогою різних протоколів зв’язку, таких як послідовні порти (наприклад, RS-232), MQTT, Bluetooth, Wi-Fi, Ethernet ZigBee, USB та інші . Якщо шлюз не має можливості розуміти дані, які надходять від об’єктів зондування за допомогою різних протоколів зв’язку, не буде ефективного зв’язку та контролю над цим середовищем.

Крім того, очікується, що сенсорні об’єкти діятимуть або автономно, або за вказівками інших об’єктів у разі виникнення події. Це вимагає генерації додаткових даних, а потім надсилання їх до відповідних об’єктів для здійснення фізичних подій. Для цього шлюзу може знадобитися підключитися до зовнішнього об’єкта через хмару, щоб отримати аналітику та статистичні дані. Це збільшує життєздатність шлюзів на рівні пристроїв в архітектурі IoT. Таким чином, необхідні захист і відмовостійкість, щоб уникнути перетворення шлюзів на єдину точку збою. Мережевий рівень в IoT-архітектурі забезпечує передачу даних від сенсорних пристроїв і шлюзів до хмарних сервісів через Інтернет. Шлюзи та деякі сенсори підтримують інтернет-протоколи, такі як IPv4, IPv6 і 6LoWPAN. Рівень підтримки служб і додатків виконує функції, подібні до транспортного і сеансового рівнів у моделі OSI. Він відповідає за контроль типу передачі, якості з'єднання та вибору порту призначення. Інакше кажучи, цей рівень спрямовує дані від сенсорів до відповідної програми на прикладному рівні для подальшої обробки. Прикладний рівень, у свою чергу, є середовищем для роботи програм, таких як ті, що показані у верхній частині рис. 1.9. До типових додатків IoT належать розумний транспорт, розумні будівлі, технології "розумного життя" тощо [12].

Функції управління в архітектурі IoT включають інструменти для віддаленого моніторингу та управління пристроями. Для реалізації віддаленого управління використовуються протоколи з'єднання, такі як SSH та VPN, що дозволяє отримувати доступ до віддалених пристроїв і виконувати операції управління, наприклад, оновлення прошивки, внесення виправлень та налаштування.

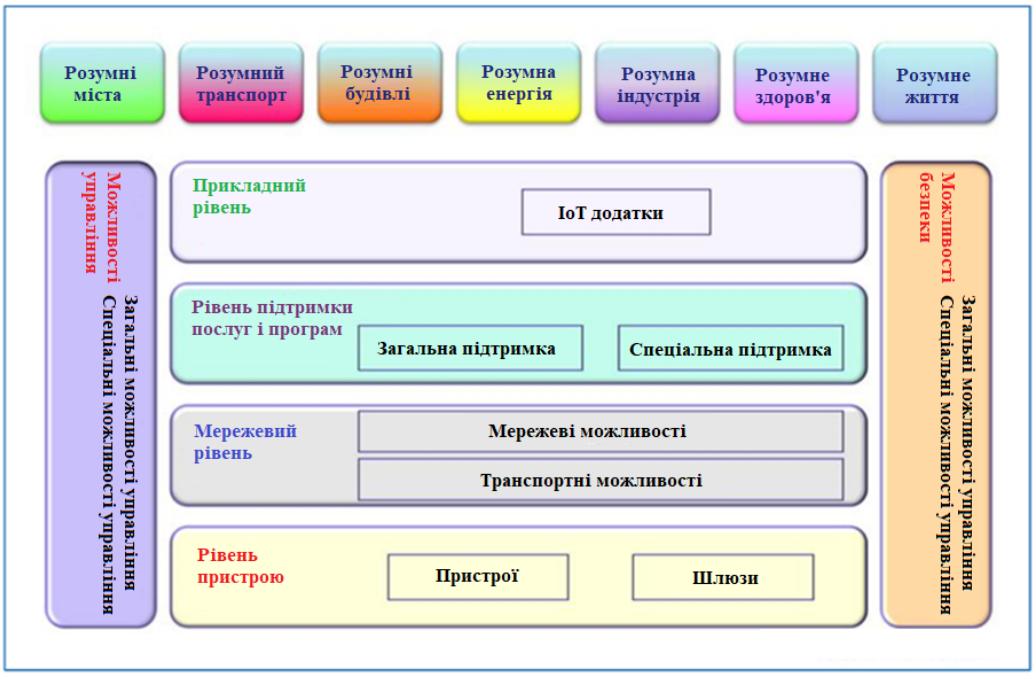

Рис. 1.9 Підходи до архітектури IoT, пропоновані IUT

Управлінські функції також охоплюють моніторинг активів, при цьому кожен елемент у мережі має свій унікальний ідентифікатор, що дозволяє зв'язати його з відповідними атрибутами. До атрибутів належать ідентифікаційний номер, виробник, дата придбання, дата введення в експлуатацію, функціональні можливості, підтримувані протоколи та інші дані. Безпекові функції стосуються інструментів і заходів, спрямованих на захист об'єктів моніторингу та інформації, яку вони генерують під час передачі даних. Це передбачає використання різних типів контрольних механізмів, з яких деякі є фізичними, а більшість – логічними. Фізичні елементи управління можуть включати самі IoT пристрої, наприклад, системи відеоспостереження, в той час як логічні елементи контролю включають механізми управління доступом, ідентифікацією та шифрування, які забезпечують захист від несанкціонованого проникнення або змінення даних та пристроїв.

1.5 Аналіз особливостей виявлення аномалій в IOT

Інтернет речей (IoT) з роками зріс експоненціально, його застосування охоплює від охорони здоров’я до промислових пристроїв. З його розвитком він

забезпечує безпрецедентний рівень підключення, як ніколи раніше. Обсяг створюваних даних також зростає експоненціально, оскільки підключається більше пристроїв. Сортувати цей величезний обсяг даних і впорядкувати їх упорядковано – складне завдання.

IoT можна класифікувати на три-, чотири-, п’яти- або семирівневу архітектуру [10], тоді як загалом чотирирівнева архітектура вважається основним компонентом IoT. Ці чотири рівні: рівень сприйняття, рівень мережі, рівень проміжного програмного забезпечення та рівень додатків [11]. Рівень сприйняття містить фізичні пристрої, такі як датчики та виконавчі механізми, які збирають дані для обробки. Мережевий рівень є комунікаційним шлюзом для рівня сприйняття та системи IoT. Рівень проміжного програмного забезпечення в якому складені дані з рівня сприйняття обробляються, зберігаються та керуються ними. Нарешті, прикладний рівень містить програми кінцевого користувача, які зберігають усі оброблені дані у значущих значеннях. Інші дослідження вважають, що більше рівнів є невід’ємною частиною архітектури IoT, наприклад, рівень безпеки, рівень керування [6], бізнес-рівень [1] і рівень середовища [8], які також можна розглядати як рівень керування.

IoT-атаки класифікуються на чотири типи: фізичні, шифрувальні, мережеві та програмні атаки [9]. У середовищі IoT було багато атак, а саме атаки переповнення буфера, атаки грубої сили, отруєння DNS, атаки ін’єкцій, атаки повторів, атаки DDoS, ін’єкції SQL, бекдори та інші. Крім того, було проведено дослідження, яке виявило, що IoT можна використовувати для сприяння насильству між інтимними партнерами, які живуть у розумному домі. Багато атак в IoT можна запобігти за допомогою механізму виявлення аномалій, який може надсилати сповіщення, коли виявляється будь-яка незвичайна поведінка. Це може допомогти запобігти атакам під час їх спроби або вказати на проблеми з функціонуванням системи, які можуть призвести до простою або збою.

Виявлення аномалій — це механізм безпеки, який розрізняє, коли поведінка системи відхиляється від нормального базового рівня. Він може бути заснований на хості (HIDS) або на основі мережі (NIDS) і є невід’ємною частиною систем

Інтернету речей, оскільки він може виявляти зміни в показаннях датчиків, аномалії мережі тощо. Системи виявлення вторгнень класифікуються на три типи: на основі сигнатур, на основі аномалій і на основі протоколу стану [12]. Крім того, методи IDS можуть бути реалізовані трьома способами: контрольованим, неконтрольованим або напівконтрольованим, які можуть бути реалізовані за допомогою ШІ, статистичного моделювання тощо. Щоб виявити аномалії, система спочатку має бути навчена тому, яка поведінка є нормальною в даній системі та як виглядає звичайна схема трафіку. Відхилення від цієї нормальності вважатиметься аномалією. Навчання системи вимагатиме великої кількості складних даних IoT із звичайною схемою мережевого трафіку та значного часу для створення профілю на основі даних IoT. Крім того, традиційні методи не корисні для виявлення нових загроз і потребують більше часу для оновлень, які можна пом’якшити із застосуванням методів ШІ, а саме машинного навчання (ML) і глибокого навчання (DL). ML — це підгалузь штучного інтелекту, яка містить алгоритми та моделі, які допомагають виконувати завдання за допомогою шаблонів навчання та зв’язків, а не бути явно запрограмованими для цього. DL є підмножиною ML, яка використовує штучні нейронні мережі, які є більш складними та можуть краще працювати з великими обсягами складних даних. Методи ML і DL можуть використовувати складні аналітичні методи для спільного використання величезних і складних даних систем IoT для формування нормальної базової лінії для мережевого трафіку пристроїв IoT [13]. Це призведе до підвищення точності, швидшого часу відгуку, рентабельності, виявлення в реальному часі тощо. У результаті методи ML і DL можуть допомогти виявити, коли система відхиляється від базової лінії. ML і DL можуть виявляти аномалії, вивчаючи зв’язки та закономірності з даних, які потім можна використовувати для розрізнення нормальної та ненормальної поведінки. Однак відмінності між методами ML і DL полягають у їхній архітектурі та складності, при цьому DL є більш складним, оскільки він має справу з нейронними мережами. DL використовує нейронні мережі для вивчення ієрархічного представлення даних, що дозволяє вивчати складні шаблони. Технології ML і DL у поєднанні з IoT забезпечують ефективне

виявлення аномалій, що дозволяє швидко знаходити та виправляти аномалії. Це посилює цілісність, надійність і ефективність систем IoT. Цю комбінацію також можна використовувати в будь-якому домені IoT, наприклад, в охороні здоров'я, промислових умовах, розумних будинках тощо.

Для навчання алгоритмів ML і DL з даними IoT необхідно сформувати набір даних, який в ідеалі має містити дані системи IoT у реальному часі. Однак через складність даних IoT набори даних попередньо формуються шляхом збору різних типів трафіку в системах IoT разом із сигнатурами атак. Ці набори даних використовуються для навчання алгоритмів ML і DL та аналізу ефективності їх різних алгоритмів. Існуючі дослідження згадують, що сформовані набори даних повинні імітувати налаштування реального світу, а також мають бути комплексними та позначеними. Для використання наборів даних у техніках ML і DL також підкреслюється важливість методів вилучення ознак, очищення даних і процедур кондиціонування. Точність набору даних до даних реального світу призведе до надійних і надійних результатів виявлення за допомогою штучного інтелекту. Зазвичай використовуваними наборами даних є набори даних IoT-23, DS2OS, Bot-IoT тощо [14].

РОЗДІЛ 2 МЕТОДОЛОГІЯ ВИЯВЛЕННЯ АНОМАЛІЙ В ІОТ

У даному розділі виконано огляд досліджень у сфері систем виявлення вторгнень. Проаналізовано різні методи, які використовуються для досягнення основних результатів, і показано, як виявлення вторгнень для IoT має свої особливі проблеми через різнорідність пристроїв, обмежені обчислювальні можливості та величезну кількість підключених пристроїв, що ускладнює вбудовування системи виявлення вторгнень. Системи виявлення вторгнень (IDS) використовуються для захисту даних, якими обмінюються кінцеві точки та процеси в інформаційній системі, від несанкціонованого доступу та модифікації.

2.1 Застосування машинного навчання для визначення аномалій

МН є одним з 10 перспективних технологій. На рис. 2.4 показано, що ML входить до чотирьох найкращих технологій, які можуть принести велику користь більшості користувачів. Він відображений кружечком який має червоний колір, щоб відобразити, що ML ще не прийнято більшістю IT-директорів – керівників інформаційних систем, оскільки для цього потрібні дорогоцінні ресурси – позначено «Н» на рис.2.1 [15].

Можна стверджувати, що комп'ютерна програма вважається навченою, якщо її здатність виконувати певні завдання покращується завдяки досвіду. Таким чином, взаємозв'язок між досвідом і ефективністю можна описати як лінійну залежність: більший обсяг досвіду призводить до покращення продуктивності. Таким чином, система, яка може приймати попередні дані в якості входу, використовувати їх для аналізу і системного виявлення закономірностей, а потім робити прогнози, є навчальною системою. Завдяки постійному процесу навчання і коригування, точність і надійність прогнозів, які генерує система, зростає. Нехай "навчання" буде важливим аспектом в машинному навчанні, проте це не є кінцевою метою. Основна мета машинного навчання полягає в розробці системи, яка здатна автоматично і точно виявляти значущі шаблони у даних.

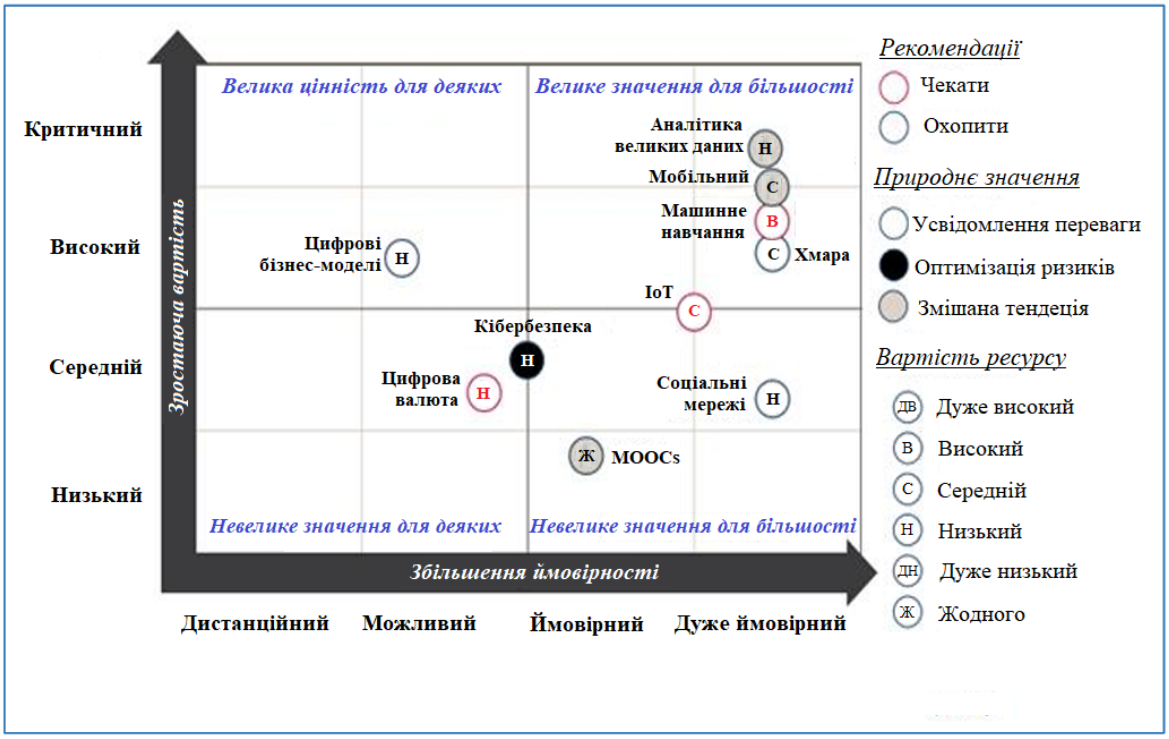

Рис. 2.1 Машинне навчання

За визначенням дослідників, "навчання" в машинному навчанні може мати кілька форм:

1. Контрольоване навчання: тут використовуються дані, що мають позначки. Викладач надає системі дані з мітками для створення бази знань, на основі якої система намагатиметься розпізнавати незмічені дані, коли вони будуть представлені. Перед початком навчання викладач повинен позначити дані, застосовуючи зрозумілі мітки, класи чи теги. Наприклад, у випадку з набором даних, що містить мільйони зображень, викладач може надати кожному зображенню короткий опис, як-от "Собака" чи "Корова". Після процесу маркування, дані передаються системі для створення її бази знань. Коли система отримує незмічені дані, такі як зображення "Лева", вона повинна вміти давати відповіді на запитання, наприклад, яка тварина відображена на цьому зображенні?

2. Неконтрольоване навчання: в даному типі навчання використовуються немарковані дані. На відміну від контрольованого навчання, де викладач знає мітки або класи для кожного елемента вхідних даних, у неконтрольованому навчанні викладач не має цієї інформації. Натомість система сама шукає мітки або класи.

Цей процес вимагає від машини проводити детальний аналіз і класифікацію, розподіляючи дані на групи з подібними характеристиками на основі атрибутів, обраних системою або викладачем. Після цього викладач може надати назву кожній групі, що дозволить системі робити прогнози для нових немаркованих даних, коли вони з'являються. Для кращого розуміння цього підходу, уявімо набір даних з мільйонами записів транзакцій. Учитель може попросити машину розділити ці записи на групи відповідно до IP-адрес, з яких виконувались транзакції. Інший викладач може вказати на необхідність групування даних за IP-адресами та часом транзакцій. Після цього викладач може географічно позначити групи та дослідити їх, щоб виявити шаблони купівельної поведінки різних груп людей відповідно до їхнього місця розташування. Ця модель буде здатна робити рекомендації користувачам на основі їхнього географічного положення, подібно до того, як Amazon пропонує товари, орієнтуючись на користувачів свого сайту.

3. Навчання з підкріпленням: у цьому типі навчання використовуються як позначені, так і немарковані дані для побудови бази знань. Цей метод базується на системі винагород, де вчитель нагороджує систему за кожен вдалий або невдалий прогноз. Винагорода слугує зворотним зв'язком, на який система спирається, коли робить подальші прогнози. З часом система формує знання на основі виконаних дій, що визначаються вхідними даними. При отриманні нового введення система намагається знайти оптимальний шлях або комбінувати кілька шляхів, щоб згенерувати прогнози і отримати винагороду. Якщо ця винагорода є кращою, ніж попередні за тією ж інформацією, то цей шлях вважається успішним. Метод навчання з підкріпленням використовується в онлайн-іграх, таких як шахи. Коли машина грає проти людей, вона записує свої ходи і реагує на дію суперника, чекаючи зворотного зв'язку від вчителя. Усі успішні кроки винагороджуються, що дозволяє машині розробити нові тактики, які в певний момент допоможуть їй обіграти людей. Унікальність цих методів навчання полягає в результатах моделі. У випадку контрольованого та неконтрольованого навчання, результати прогнозів є чітко формальними з оцінкою ймовірності. Модель генерує прогнози в межах від 0 до 1: чим ближче значення до 1, тим впевненіше система. У навчанні з

підкріпленням система навпаки створює прогнози з вірогідністю 1, чекаючи реакції вчителя, щоб дізнатися, чи були ці прогнози правильними.

Говорячи про визначення машинного навчання (ML) відповідно до досвіду Мітчелла, важливо зазначити, що для навчання використано вхідні дані, які можуть бути як позначеними, так і непозначеними. Непізнані дані — це необроблені інформаційні набори, які не містять жодних тегів, і їх значно легше отримати в порівнянні з позначеними даними. Проте обробка непозначених даних вимагає більше часу та ресурсів для моделей ML, а також для створення прогнозів. З іншого боку, позначені дані — це вже оброблені набори, що відповідають відповідним тегам. Отримати дані з мітками складніше, ніж непозначені, і звичайно це не безкоштовно, після чого в процесі беруть участь люди, які оцінюють та мітять ці дані [16].

Алгоритми ML. Алгоритм — це набір правил, яких слід дотримуватися в певному порядку, щоб вирішити проблему. У ML є багато алгоритмів, які можуть мати схожі функції. Алгоритми ML можна розділити на такі групи:

1. Алгоритми кластеризації: ці алгоритми пов’язані з ідентифікацією структур або шаблонів у даних, щоб організувати їх у групи. Кожен предмет у групі був би схожий один на одного за структурою до максимального рівня. Нижче наведені найпопулярніші алгоритми кластеризації: a. K-Means; b. К-медіани; в. Ієрархічна кластеризація; d. Максимізація очікувань (EM)

2. Алгоритми регресії: ці алгоритми пов’язані зі статистичним аналізом даних з метою оцінки зв’язку між змінними. Ці алгоритми зосереджені на зв’язку між однією залежною змінною та однією чи кількома незалежними змінними. Нижче наведено найпопулярніші алгоритми регресії: a. Лінійна регресія; b. Логістична регресія; в. Звичайна регресія найменших квадратів (OLSR); d. Поетапна регресія.

3. Алгоритми дерева рішень: ці алгоритми пов’язані з побудовою моделей прийняття рішень на основі фактичних значень атрибутів у даних. Ці алгоритми працюють шляхом навчання системи на даних, які використовуються в класифікації та регресії. Алгоритми шукають структуру дерева (тобто розгалуження), доки не будуть прийняті рішення. Нижче наведено найпопулярніші

алгоритми дерева рішень: a. C4.5 і C5.0; b. M5; в. Дерево класифікації та регресії (CART); d. Пень рішення.

4. Алгоритми Байєса: ці алгоритми безпосередньо пов’язані із застосуванням теореми Байєса в класифікації та регресії. Коротше кажучи, теорема Байєса вивчає умовну ймовірність кожної можливої причини даного результату. Нижче наведено найпопулярніші алгоритми Байєса: a. Наївний Байєс; b. Багаточлен наївного Байєса; в. Гаусівський наїв; d. Байєсова мережа

5. Алгоритми штучної нейронної мережі: ці алгоритми пов’язані із застосуванням алгоритмів регресії та класифікації подібно до біологічних нейронних мереж людини. Нижче наведено приклади алгоритмів штучної нейронної мережі: a. Зворотне поширення; b. Перцептрон; c. Мережа Мережа Хопфілда

На рис. 2.2 показано базовий процес алгоритмів ML, де вхідні дані є позначеними або не позначеними даними IoT, а вихідними є система сповіщень, яка зазначає, чи є дані аномальними чи нормальними [17].

Рис.2.2 Основний процес алгоритму машинного навчання, навченого за допомогою даних IoT для виявлення аномалії

Дослідження виявлення аномалій на основі машинного навчання поділяються, по-перше, на статті, в яких обговорюється виявлення аномалій і IDS, що включає виявлення аномалій і атаки, а по-друге, на атаки, які відбуваються в мережах IoT. Атаки в Інтернеті речей є аномаліями в контексті виявлення аномалій, оскільки атаки вимагають від системи зображення незвичайної поведінки, щоб бути успішними. Трафік може бути аномальним через зловмисне корисне навантаження, аномалії поведінки, незвичний мережевий трафік тощо. Тому було проведено комплексне дослідження різних аспектів виявлення аномалій за

допомогою ML. Підсумок усіх досліджень та їх результати для виявлення аномалій за допомогою ML представлено в табл. 2.1 , а виявлення аномалій на основі атаки представлено в табл.2.2.

Таблиця 2.1

Виявлення аномалій за допомогою ML

Проблему вирішено	Набір даних	Пропоноване рішення	Отримані результати	Переваги	Недоліки	рік
Виявлення аномалій в IoT за допомогою ML	Дані часових рядів, NSL-KDD	Порівняно кілька класифікаторів ML, таких як KNN і DTs	DTs і лінійний дискримінантний аналіз досягли 80% точності з даними, що не є часовими рядами	Використовуються як дані часових рядів, так і дані не часових рядів	Невеликий розмір набору даних	2018 рік
Виявлення аномалій і атак в мережах IoT	Набір даних від Kaggle	Порівнювали різні моделі машинного навчання для прогнозування атак і аномалій у системах IoT	RF та ANN перевершили DT з точністю 99,4%.	Моделюйте краще та швидше, ніж інші техніки	Обмеження набору даних і обчислювальна складність	2019 рік
Виявлення аномалій в IoT за допомогою ML	IoT-23	Досліджували моделі ML, порівнювали алгоритми та оцінювали їх продуктивність за допомогою метрик	РЧ перевершив інші з показником точності 99,9%	Точно визначені аномалії в мережевому трафіку	Потрібне тестування з більшою кількістю наборів даних, крім IoT-23	2020 рік
Використання моделей на основі ML для виявлення аномалій в IoT	ToN-IoT та BoT-IoT	Запропонував структуру на основі Hadoop із використанням класифікаторів KNN, SVM і NB ML	Точність становила 90% з ToN-IoT і 99% з BoT-IoT	Висока точність і низький рівень помилкових позитивних результатів	Потребує тестування з більшими наборами даних і складнішими алгоритмами машинного навчання	2021 рік
иявлення аномальної активності в системах IoT за допомогою ML	IoT-23	Оцінка методів градієнтного посилення та екстремального градієнтного посилення (XGBoost) за допомогою набору даних IoT-23	XGBoost мав високий рівень точності 99,98%	XGBoost може підвищити безпеку системи IoT	Потрібне більше тестування на більших і реальних наборах даних	2022 рік
Виявлення аномалій в IoT за допомогою ML	CoAP-IoT	Представлено новий набір даних CoAP-IoT і підтверджено його за допомогою навчання під наглядом	RF, SVM і DT показали найкращі результати із середньою точністю 0,9	Створив новий набір даних і підтвердив його	Потребує додаткового тестування з використанням різних наборів даних і реальних систем Інтернету речей	2023 рік

Таблиця 2.2

Виявлення аномалій на основі атаки

Проблему вирішено	Набір даних	Пропоноване рішення	Отримані результати	Переваги	Недоліки	рік
DDoS-атаки в IoT	дані PCAP	Використані класифікатори ML QDA, SVM, KNN, NV, DT і RF	DPAE перевершив інші моделі	Класифікатор на основі SDN	Неправильна класифікація законного трафіку та затримки виявлення	2018 рік
DDoS у сценаріях IIoT	UNSW NB15	Запропоноване об’єднане навчання для виявлення DDoS в IIoT	Низький час відгуку пом'якшення з високою точністю пом'якшення	Висока точність і малий час відгуку	Потрібні додаткові тести для реалізації в реальних налаштуваннях Інтернету речей	2020 рік
ловмисний бот-трафік IoT у мережах IoT	Бот-IoT	Запропонував нову метрику під назвою CorrAUC на основі метрики AUC	Модель була ефективною і досягла 96% точності	Висока точність виявлення шкідливого трафіку	Не масштабується та потребує додаткових навчальних даних	2021 рік
Ідентифікація зловмисного ПЗ IoT	IoT-23, LITNET-2020 і NetML-2020	Використовував алгоритми ML і DL для наборів даних для виявлення зловмисного програмного забезпечення в наборах даних	RF досяг найвищого результату точності 96%	Виявляє високу точність у класифікації шкідливих програм	Управління великими наборами даних	2022 рік
Кібератаки та IDS в мережах IoT	IoT-23 і вторгнення в мережу IoT	Порівняно різні моделі ML — RF, DT, NB, MLP і KNN	RF і DT показали найкращі результати з точністю 99,9% кожен	Висока точність у класифікації шкідливих дій	Потрібні більші набори даних і підвищена точність моделей	2023 рік
Виявлення шкідливих програм в IoT	UNSW-NB15	Використані моделі ML — LR, KNN, DT, ET, RF і MLP	ET досяг 99,98% точності	Використовується великий і різноманітний набір даних	Важко виявити атаку нульового дня	2024 рік

Глибоке навчання — виявлення аномалій мережі IoT. Глибоке навчання (DL) передбачає використання архітектури нейронної мережі для виявлення аномалій у системі. Нейронні мережі створені для імітації людського мозку. де кожен шар обробляє вхідні дані [18]. Завдяки використанню нейронних мереж DL може обробляти великі та складні дані для вилучення відповідної інформації. DL можна використовувати як контрольоване, неконтрольоване або напівконтрольоване навчання. Різні типи моделей DL включають згорткові

нейронні мережі (CNN), повторювані нейронні мережі (RNN), часові згорткові мережі (TCN), автокодери, генеративні суперницькі мережі (GAN), мережі глибоких переконань (DBN), довготривалу короткочасну пам’ять (LSTM).) Нейронні мережі тощо. Порівняно з моделями ML, згадується, що моделі DL краще виявляють аномалії, які раніше не виявлялися, і вони більш здатні аналізувати великі та складні набори даних, що робить їх придатними для IoT. Отже, вважають, що DL-моделі краще підходять для підтримки безпеки та конфіденційності в системах IoT [19]. Окрім виявлення аномалій, інші переваги моделей DL включають прогнозний аналіз, автоматизацію, підвищену точність, масштабованість, ефективність тощо. На рис.2.3 показано базовий процес роботи алгоритму на основі DL, де вхідними даними є дані із середовищ IoT, а виходом є двійкова система сповіщень, яка інформує користувача про те, чи є дані аномальними чи нормальними.

Рис.2.3 Основний процес алгоритму глибокого навчання, який використовується для виявлення аномалій в IoT

Виявлення аномалій за допомогою DL класифікується на виявлення аномалій в IoT, яке представлено в узагальненому форматі в табл.2.3 , і виявлення атак, яке представлено в табл.2.4 . Напади в цьому дослідженні вважаються аномаліями; однак ці два розділені систематично, щоб класифікувати їх.

Таблиця 2.3

Виявлення аномалій за допомогою DL

Проблему вирішено	Набір даних	Пропоноване рішення	Отримані результати	Переваги	Недоліки	рік
Виявлення аномалій на основі DL	BoT-IoT	Запропонована модель VCDL	Досягнуто 99,7% точності	Перевершив інші моделі	Проблеми класового дисбалансу	2020 рік
Виявлення аномалій в IoT на основі DL	BoT-IoT, MQTT-IoT-IDS2020, IoT-23, IoT-DS-1 та IoT-DS-2	Запропоновані моделі CNN1D, CNN2D і CNN3D	Усі моделі досягли точності >99% для всіх наборів даних	Запропоновані моделі DL перевершили інші моделі	Обмежені набори даних і відсутність фактичного тестування	2022 рік
IDS в IIoT	WUSTL-IIOT-2021	Використані моделі DL з даними мережевого потоку для IDS	Отримано 99% рейтингу точності	Успішно впорався з дисбалансом класів у наборі даних	Потрібне більше тестування з більш різноманітними наборами даних	2023 рік
Виявлення аномалій об'єднаним DL	UNSW-NB15	FDQN використовується в наборі даних для виявлення аномалій	Кращі показники використання ресурсів і точності виявлення	Масштабований, універсальний і перевершує інші моделі	Точні значення метрик не називаються	2023 рік
IDS в IoT з моделями на основі DL	ToN_IoT, CICIDS2017 і SWaT	Запропоновано стек-ансамбль моделей DL під назвою DIS-IoT	Показник точності з ToN_IoT становив 99,6%, з CICIDS2017 – 98,7%, а з SWaT – 99,7%	Перевершив інші моделі за всіма показниками	Потребує тестування на реальних пристроях IoT	2024 рік

Таблиця 2.4

Виявлення атак за допомогою DL

Проблему вирішено	Набір даних	Пропоноване рішення	Отримані результати	Переваги	Недоліки	рік
IDS для виявлення атак	KDD99	RBM використовується для виявлення	Було досягнуто рівня точності 94%.	Здатність моделей DL виявляти атаку	Комплексні результати не згадуються	2018 рік
Виявлення атак в IoT	IoT-23	Гібридна модель DL CNN і LSTM	Досягнута точність виявлення 96%	Покращена точність і ефективність	Потрібне тестування з більшою кількістю наборів даних	2021 рік
Виявлення DDoS і DoS атак в IoT	Зібрані дані та N-BaIoT	Модель глибокого ансамблевого навчання DeL-IoT	Перевершує методи ML із рівнем виявлення 99,8%.	Забезпечує точність і масштабованість	Потрібні додаткові тести з різноманітними наборами даних	2021 рік

Продовження таблиці 2.4

Виявлення атак за допомогою DL

Проблему вирішено	Набір даних	Пропоноване рішення	Отримані результати	Переваги	Недоліки	рік
Виявлення кібератак за допомогою DL	CIC IoT 2022	Для перевірки набору даних використовувалися FFNN, LSTM і RandNN	Оцінка точності FFNN становила 99,93%, LSTM – 99,7%, RandNN – 96,42%	Універсальні функції вилучення та класифікації	Потрібна оптимізація з більш різноманітними наборами даних	2023 рік

Зрештою, використання алгоритму на основі DL є перспективним для точного виявлення аномалій. Однак проблема з алгоритмами на основі DL полягає в тому, що вони потребують великих і високоякісних наборів даних, є важкими для обчислень і потребують часу для навчання. Крім того, складність алгоритмів DL ускладнює визначення причини або шляху процесу прийняття рішень.

Для досліджень, що базуються на ML, найчастіше цитована модель має найвищу точність – це модель Random Forest (RF), яка цитувалася близько дванадцяти разів як найвища точність [20]. Найменше згадувані моделі високої точності – це моделі ANN, GBM і RT. Різноманітне використання RF протягом багатьох років робить його точною моделлю ML для виявлення аномалій і атак. Більшість досліджень мають загальний недолік, який полягає в необхідності більш різноманітних наборів даних для перевірки запропонованих моделей [16]. Це супроводжується недоліком, що моделі можуть бути важкими для обчислень для систем IoT. На рис.2.4 наведено підсумок наборів даних, які найчастіше використовуються в дослідженнях, які розглядаються для цього дослідження. Згідно з малюнком 4 серед усіх наборів даних UNSW-NB 15 та IoT-23 є наборами даних, які найчастіше використовуються як для тестування моделі ML, так і для моделі DL.

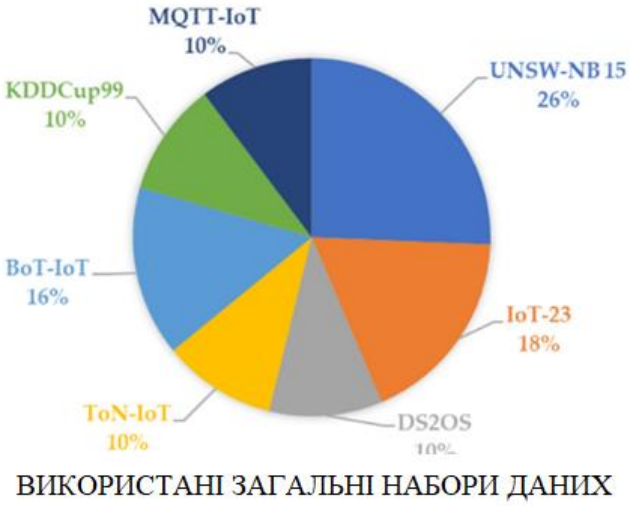

Рис.2.4. Найпоширеніші набори даних, які використовуються в наукових роботах, проаналізованих у цьому дослідженні

Для досліджень на основі DL модель, яку найчастіше цитують із найвищою точністю, це нейронна мережа довготривалої короткочасної пам'яті (LSTM) та її гібрид [19], які з'являлися близько семи разів, за якою слідує звичайна нейронна мережа (CNN) та її гібрид, який з'являвся приблизно шість разів, за якими слідує Deep Neural Network (DNN). Отже, з літератури моделі LSTM, CNN і DNN виявляються найточнішими моделями DL у виявленні аномалій і атак. Найпоширенішими недоліками моделей на основі DL є потреба в більш різноманітних наборах даних і більших наборах даних, а також обчислювальна складність моделей [21].

2.2 Аналіз методів виявлення вторгнень

До проблеми побудови ефективних моделей класифікації для IDS зазвичай підходять з використанням кількох алгоритмів кластеризації даних і класифікації. Дослідження [22] представляє класичний гібридний підхід, ближчий до того, що ми реалізуємо в цій дипломній роботі. Пропонує гібридну систему виявлення вторгнень на основі Support Vector Machine (SVM) і C5.0. Автор стверджує, що використання такої комбінації алгоритмів підвищить точність виявлення вторгнень у порівнянні з використанням SVM і C5.0 окремо. Дослідження [23] пропонує

використовувати SVM щоб здійснювати класифікацію інформації у класи. Враховуючи навчальні дані, SVM має можливість відображати дані у багатовимірному просторі ознак за допомогою нелінійних відтворень приклад якого продемонстровано на рис.2.5. Обчислюючи роздільну гіперплощину для заданого навчального набору даних, можна визначити опорні вектори, які згодом застосовуються для створення теоретичного запасу. Цей запас дозволяє SVM прогнозувати, до якого класу або категорії належить нова вибірка даних. У дослідженні [24] пропонується поєднати SVM з алгоритмом C5.0, що є популярною реалізацією дерева рішень. Алгоритм C5.0 будує дерево, яке починається з кореневого вузла та містить кілька дочірніх вузлів. Кожен вузол відповідає за аналіз окремих атрибутів чи кортежів у вхідних даних. Усі вузли підпорядковуються кореневому, який використовує інформаційну теорію для формування рішень.

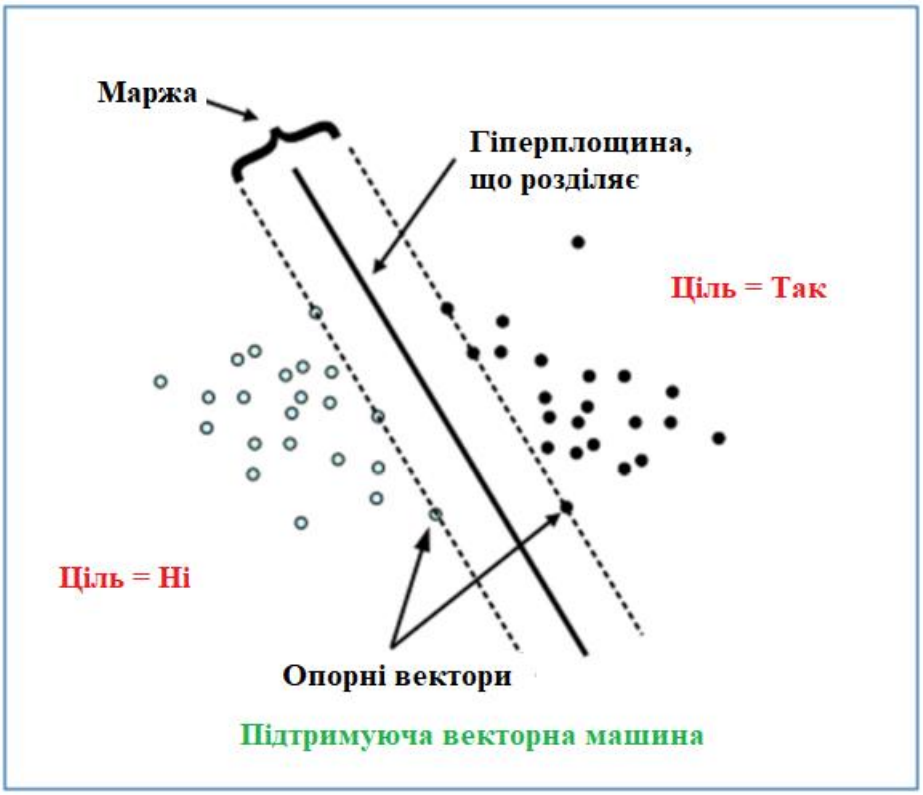

Рис.2.5 Основні концепції SVM

У дослідженні [25] для перевірки запропонованої методики виявлення вторгнень використовується набір даних KDD Cup. Ці дані були поділені на тренувальну (75%) і тестову (25%) вибірки. Модель Голмана спершу обробляє тренувальні дані за допомогою C5.0, який додає до вихідного набору даних інформацію про вузли, формуючи нову версію навчальної вибірки. Отриманий набір передається до SVM для оцінювання. Порівнюючи гібридну модель із традиційною SVM, автори [24] виявили, що середня точність моделі на основі C5.0 і SVM становить 99,96%, тоді як окрема SVM досягла точності 99,78%.

У недавньому дослідженні [25] акцентується увага на тому, що постійне зростання атак на критичну інфраструктуру та недостатня ефективність існуючих систем виявлення пояснюються складністю розпізнавання нових, «невідомих» атак. Таким чином, система, запропонована в [25], може бути обмеженою, оскільки під час навчання покладається на дані з мітками, які обробляються C5.0.

Натомість у роботі [26] представлено нову модель виявлення вторгнень, що використовує MapReduce для обробки великих обсягів структурованих і неструктурованих даних у вигляді пар ключ/значення. Процес створення цих пар базується на поєднанні кластеризації Fuzzy C-Means (FCM) і SVM для класифікації. Перевага FCM полягає в тому, що він дозволяє точкам даних належати до кількох кластерів, на відміну від нечіткого CM, де кожна точка може входити лише до одного кластеру.

У дослідженні [27] пропонується гібридна система виявлення вторгнень, яка використовує методи інтелектуального аналізу даних, зокрема K-середні для кластеризації та Наївний Байєс для класифікації даних.

Напади, здійснені у [22] модель призначена для виявлення є; DoS, Probe, U2R і R2L. Ці класи стосуються типів атак, які існують у наборі даних KDD cup 99. Класи атак пояснюються наступним чином:

1. DoS – відмова в обслуговуванні; це атака, яка виводить систему з ладу, використовуючи ресурси цієї системи, включаючи обробку, зберігання та ресурс пам’яті.

2. R2L– віддалене до локального; це атака, спрямована на отримання доступу до локальної машини віддалено з використанням будь-якої існуючої вразливості в системі.

3. U2R – користувач для рутування; це атака, яка має на меті посилити доступ до системи від звичайних привілеїв користувача до прав адміністратора.

4. Зондування – що відноситься до сканування дій, які атака буде виконувати для збору даних про об’єкт до того, як атакувати її.

У роботі [25] спочатку був застосований метод K-середніх за допомогою якого виконується кластеризація інформації на три групи: DoS, Probe і «інші». До кластеру «інші» включаються R2L, U2R і звичайні дані. Хоча розрізнити атаки DoS і Probe між собою і також від іншого типу атак відносно просто, відокремлення R2L, U2R і звичайних даних є значно складнішим завданням через їхню схожість. Для вирішення цієї проблеми [25] пропонують використовувати класифікатор Naive Bayes, який здатний класифікувати атаки та звичайний трафік. Цей класифікатор характеризується як потужний алгоритм навчання, що визначає критерії класифікації на основі взаємозв’язків між залежними та незалежними змінними. Однак дослідники [25] не оприлюднили результати тестування запропонованої моделі, що унеможливлює оцінку її точності.

2.3 Виявлення вторгнень для IoT

Огляд моделей виявлення вторгнень, описаних у [25], показує, що методи машинного навчання активно застосовуються для кластеризації та класифікації даних з метою виявлення мережевих атак, і вони широко визнані ефективними. Автори також підкреслюють, що пристрої IoT не менш вразливі до таких атак. Більше того, дані, якими обмінюються IoT-пристрої, можуть бути не менш, а іноді й більш важливими, ніж дані від пристроїв, що не належать до IoT. Однак обмежені обчислювальні ресурси IoT-пристроїв ускладнюють інтеграцію надійних рішень для забезпечення безпеки.

У роботі [20] пропонується легка система виявлення вторгнень як альтернатива більш ресурсомістким IDS, які використовують машинне навчання. Ця система націлена на ідентифікацію вузлів у IoT-середовищі, які мають кілька ідентифікаторів. Модель, описана в [16], оптимізована для бездротових мереж і базується на аналізі потужності сигналу, а також використанні секретних ключів та ідентифікаторів для виявлення вузлів із множинними ідентифікаторами, що намагаються підключитися до прикордонного маршрутизатора IPv6.

Варто зауважити, що модель, запропонована у [23], більше схожа на систему патрулювання, а не на повноцінну систему виявлення аномалій, оскільки вона не враховує аналіз корисного навантаження, яке вузли можуть передавати до центру керування.

У [25] дослідники представили новий підхід до виявлення вторгнень у IoT, орієнтований на виявлення аномалій. Вони окреслили ключові проблеми IoT-середовища, які обмежують ефективність стандартних методів виявлення вторгнень, розроблених для систем, що не належать до IoT. По-перше, IoT є вразливим до численних атак через відкрите розгортання і обмежені ресурси. По-друге, IoT-системи зазвичай складаються з різнорідних пристроїв від різних постачальників, які часто не мають єдиних стандартів або протоколів. Запропонована схема складається з двох етапів для вирішення цих викликів:

1. Застосування алгоритму для виявлення аномалій з метою ідентифікації відхилень у даних на рівні сприйняття.

2. Використання схеми розподіленого виявлення вторгнень для розмежування аномалій та зловмисних атак на рівні сегмента мережі, без застосування централізованої системи виявлення вторгнень.

Припускаючи, що кожен пристрій IoT оснащений терміналом, який передає необроблені дані до центральної точки збору для подальшого аналізу, ці дані називаються даними рівня сприйняття. Такі дані можуть бути різними за природою, вологими або будь-якими іншими, що датчик пристрою IoT може збирати та передавати. Дослідники роботи [22] вважають, що аномалії можуть бути виявлені

на цьому етапі шляхом аналізу патернів даних та передачі повідомлень про аномалії до системи розподіленого виявлення вторгнень на наступному етапі.

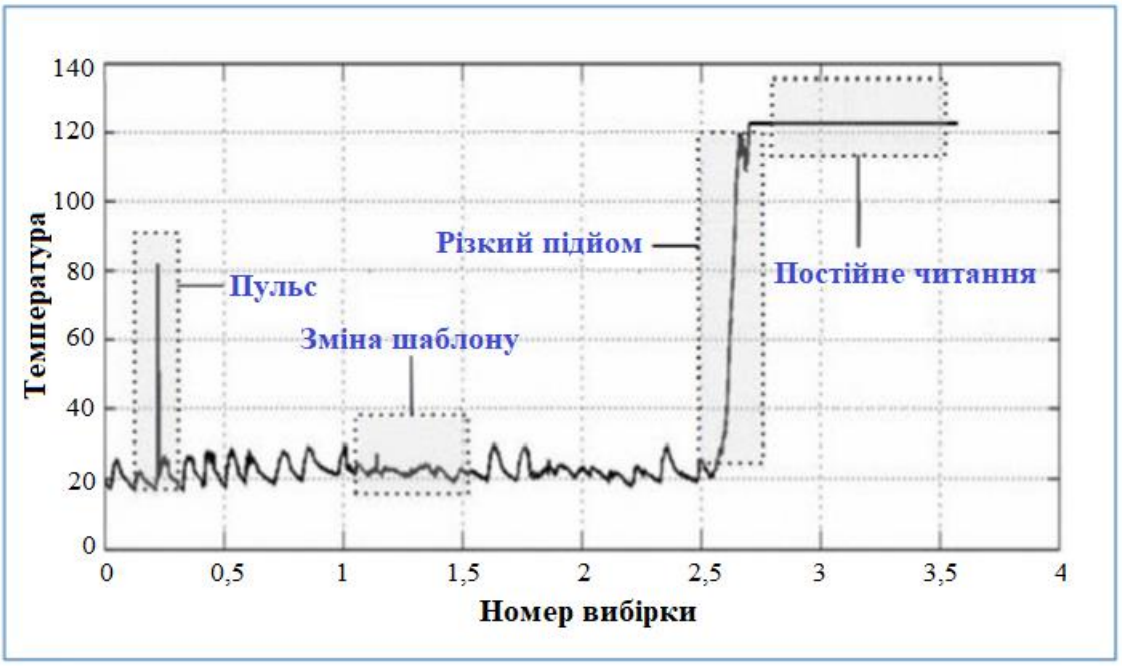

Рис. 2.6 Приклад аномалій у даних

На рис.2.6 показано зразок даних про температуру, зібраних датчиком. Стадія покладається на аналіз такої вибірки даних, щоб мати змогу виявити аномалії на базі атрибутів, які застосовуються для ідентифікації атак. Визначено 5 ознак аномалії;

1. Пульс: раптове припинення даних на дуже короткий час.

2. Різкий підйом: раптове розрив даних, що залишається протягом тривалого часу. 3. Різкий спад: раптове падіння даних, яке залишається протягом тривалого часу.

4. Постійне читання: постійне читання протягом певного часу.

5. Зміна шаблону: дані за певний період часу подібні до звичайних шаблонів. Однак може бути різниця в середньому та/або дисперсії.

Щоб спостерігати за цією діяльністю у зразках даних, використали аналітичний алгоритм під назвою SAX для розділення зразків даних на основі часу до вікон. Вікно буде містити дані за певний проміжок часу. SAX використовує 5

атрибутів аномалій, щоб порівнювати вікна одне з одним, витягувати відповідні аномалії та звітувати на етапі 2, використовуючи такий кадр:

<*terminalid, anomaly_attri, property_attri, timestamp*>

Terminalia – це ідентифікатор терміналу; об'єкт IoT на рівні сприйняття. Anomaly_attri — це атрибут аномалії, як пояснено вище, property_attri — це фактичне значення або дані, якими можуть бути напруга, температура, освітленість, вологість тощо, а позначка часу — це час, коли було виявлено аномалію. У роботі [22] використовують неконтрольований алгоритм інтелектуального аналізу даних для виявлення нормальної моделі. Хоча дослідники не розкрили, які алгоритми неконтрольованого інтелектуального аналізу даних використовувалися, можна припустити, що використовувалися кластеризація з подальшою ідентифікацією алгоритмів класифікації. Перший буде кластеризувати дані на основі подібності в невідомі групи, а другий буде класифікувати дані на нормальні та аномальні групи. Ідентифікація нормальних шаблонів у термінальних даних полегшить глибоке використання 5 аномалій, пов’язаних із подальшою класифікацією аномалій у групі аномальних даних.

Після виявлення аномалій і повідомлення про них системам виявлення вторгнень, можна класифікувати три типи активності: забруднення даних, зниження якості обслуговування та відмова в обслуговуванні. На другому етапі [22] запропоновано впровадити систему виявлення вторгнень на різних рівнях: рівні сприйняття даних, рівні розподілу даних (тобто доступу до мережі) та рівні додатків, а також забезпечити взаємодію між ідентифікаторами на цих рівнях для обміну правилами виявлення. Для оцінки своєї системи [22] використали набір даних Intel Lab Project, який складається з офісу, оснащеного датчиками для збору інформації про температуру, вологість, освітленість та напругу кожні 30 секунд, з подальшою передачею даних у центральний пункт збору. Цей набір даних містить 2,3 мільйона вимірів. На жаль, на сьогодні немає доступних загальнодоступних наборів даних, що містять анотовані атаки, тому [22] вручну додали випадкові

аномалії до показників і здійснили оцінку на основі цих змін. Інформація про точність системи не надана. Дослідники також вказують на додаткові причини, чому традиційні схеми виявлення вторгнень не є ефективними для IoT. Ці схеми є складними та обчислювально затратними, що робить їх непридатними для безпосереднього застосування в IoT. Крім того, методи виявлення аномалій на основі моделей і подібності мають проблеми з дизайном, реалізацією та високим рівнем помилкових спрацьовувань. Тому пропонується використання простого алгоритму для виявлення аномалій в IoT. У роботі [13] застосовуються евклідова відстань і коефіцієнт Жаккара для виявлення аномалій у даних. Ці математичні операції дозволяють вимірювати подібність між двома наборами даних.

У контексті IoT вважається, що кожен датчик генерує кілька записів даних. Визначивши відстань між двома записами, отриманими від одного датчика, можна оцінити їх схожість. Окрім того, знаючи звичайний стан запису, дослідники запропонували використовувати гіпотезу або порогове значення, засноване на коефіцієнті Жаккара:

При 0,8 < коефіцієнт Жаккара <=1; нормально

У разі коефіцієнта Жаккара < 0,8; аномальний

Тобто, якщо схожість між двома записами даних становить менше 0,8, то відхилення дійсно існує. Якщо подібність виявилася більше 0,8, то дані в нормі. Досліджники використовують ковзне вікно як спосіб підтримувати статистичну інформацію як потік даних. Він використовував неконтрольовані методи, щоб знайти нормальний шаблон у даних. На жаль, не розкривають жодної інформації щодо того, який набір даних вони використовували, а також результату, який вони отримали. Дослідники [16] підходять до виявлення аномалій IoT під іншим кутом зору, зосереджуючись на IoT у хмарі, що здається надзвичайно цікавим підходом через вибух хмарних обчислень і технологій великих даних. По-перше, обсяг і кількість даних, які передаються в хмару з пристроїв IoT, є значними. До 2024 року очікується підключення близько 30 мільярдів пристроїв IoT до Інтернету. По-друге, нерівномірність підключення та різноманітність трафіку пристроїв IoT

ускладнюють процес виявлення аномалій в цій сфері. Автори надають огляд методів виявлення аномалій в IoT, підсумовуючи їх таким чином:

1. Розподілені підходи: у цьому підході пристрої спільно працюють над виявленням аномалій. В основі таких методів лежать схеми виявлення на основі правил, включаючи кооперативне виявлення, статистичне виявлення аномалій та методи машинного навчання, як-от машини опорних векторів. Основні переваги розподілених підходів — це швидкість і можливість масштабування.

2.Централізовані підходи: у порівнянні з розподіленими, централізовані методи є простішими в реалізації та потребують менших обчислювальних ресурсів від пристроїв IoT. Поширеними є алгоритми на основі репутації та евристичні методи ранжування для виявлення шкідливих вузлів у бездротових сенсорних мережах WSN.

3.Ієрархічні підходи: оскільки Інтернет є ієрархічною мережею, цей підхід є оптимальним для великих мереж. Методи, як-от кластеризація, статистичне виявлення аномалій і марковські моделі, є популярними для виявлення аномалій в IoT.

4.Автономні підходи: цей метод дозволяє кожному вузлу або пристрою IoT самостійно виявляти атаки, зберігаючи короткочасну динамічну статистику подій.

Вищезгадані підходи орієнтовані на забезпечення безпеки пристроїв IoT з точки зору мережевих загроз. Тому вони можуть бути використані для виявлення мережевих атак, таких як підроблені вузли, фальшиві MAC-адреси, неправильна маршрутизація, а також атаки типу DoS і DDoS. Дослідники також пропонують нову класифікацію методів виявлення атак, поділяючи їх на три основні типи:

1. Статистичний

2. Інтелектуальний аналіз даних

3. Штучний інтелект ШІ.

Усі ці методи вимагають або вивчення попередніх зразків даних, або статистичного аналізу даних. Конечні автомати, кластеризація, мови опису та експертні системи є широко використовуваними методами інтелектуального

аналізу даних. Нечіткі, генетичні алгоритми, байєсівські, марковські та нейронні мережі широко використовуються в ШІ.

У табл.2.5 подано підсумок відповідної роботи. Спосіб підійти до таблиці полягає в усвідомленні того, що ті самі методи виявлення аномалій у випадку мережі, що не є IoT, можуть використовуватися для тієї самої мети в мережі IoT, приділяючи особливу увагу унікальним характеристикам IoT – різнорідні пристрої, неоднорідні дані, обмежені обчислювальні ресурси, які диктують використання полегшених версій цих методів для знаходження аномалій.

Таблиця 2.5

Таблиця суміжних робіт

Метод/модель	Точність	
Підтримка Vector Machine (SVM) і C5.0 для виявлення аномалій.	99,78%	Non-IoT
Модель виявлення вторгнень на основі аналізу великих даних.	Не повідомляється	
Гібридна система виявлення вторгнень із використанням двох методів інтелектуального аналізу даних; Кмеанс і Наївний Байєс.	Не повідомляється	
Штучна нейронна мережа системи виявлення вторгнень для атак IoT.	99,4 %	IoT-based
Дворівнева модель класифікації для виявлення аномалій в магістральних мережах IoT	84,82%	
Неконтрольована класифікація нормальних і ненормальних аномалій в IoT	Не повідомляється	
Легке неконтрольоване виявлення аномалій за допомогою евклідової відстані та коефіцієнта Жаккрада в IoT	Не повідомляється	

РОЗДІЛ 3 АНАЛІЗ ТА РЕЗУЛЬТАТИ ВПРОВАДЖЕННЯ МОДЕЛІ ВИЯВЛЕННЯ АНОМАЛІЙ В IOT

У цьому розділі описується методологія, спосіб, за допомогою якого модель виявлення аномалій в IoT використовує комбінацію двох алгоритмів. Алгоритму інверсної вагової кластеризації (IWC) і дерева рішень (DT). Він містить опис даних, які будуть використовуватися при розробці та тестуванні моделі, а також опис того, як IWC і DT працюють у кластеризації та класифікації даних. Він також описує критерії оцінки; критерії, за якими можна виміряти точність моделі у виявленні аномалій в IoT.

Описано впровадження досліджуваної моделі для виявлення аномалій в IoT. Модель реалізовано за допомогою MATLAB. Реалізація моделі складається з трьох етапів, а саме попередня обробка вхідних даних, кластеризація, класифікація та тестування. Попередня обробка вхідних даних є етапом кваліфікації набору даних Intel Labe IoT для використання в якості вхідних даних для процесу виявлення аномалій (ADP). Кластеризація та класифікація є основними підпроцесами ADP і моделі загалом. Кластеризація розділяє вхідні дані на дві групи даних, які мають майже схожу структуру та характеристики за допомогою алгоритму Invers Weight Clustering (IWC). Підпроцес класифікації використовує алгоритми C4.5 для побудови дерева рішень (DT), яке використовується для класифікації даних на нормальні та ненормальні. Представлені результати запуску запропонованої моделі для виявлення аномалій в IoT. MATLAB використовується для виконання тесту з використанням частини некласифікованих вхідних даних та отримання результатів.

3.1 Дизайн моделі та вхідні дані для виявлення вторгнень в IoT

Дизайн моделі. Метою запропонованої моделі виявлення аномалій в IoT є виявлення аномалій у даних, якими обмінюються пристрої IoT на прикладному рівні IoT, архітектура якого описана в попередньому розділі. Для цього модель

вимагає розробки процесу, який приймає дані з пристроїв IoT як вхідні дані, систематично обробляє вхідні дані та створює двократні прогнози; «нормальний» або «ненормальний». На рис.3.1 представлено високий рівень моделі [11].

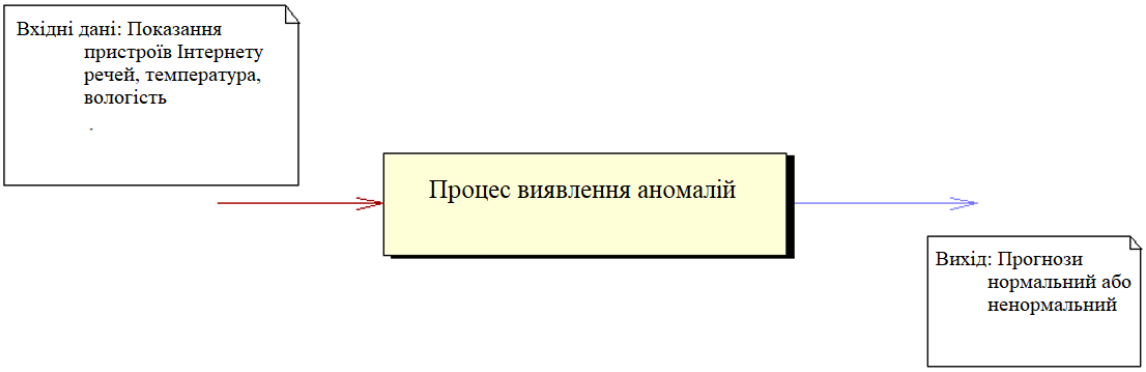

Рис.3.1 Процес виявлення аномалій

Процес включає три основні етапи:

1. Попередня обробка вхідних даних: на цьому етапі відбувається форматування вхідної інформації таким чином, щоб спростити її подальшу обробку за допомогою ADP.

2.Обробка: цей етап полягає в кластеризації вхідних даних за допомогою IWC та побудові DT для здійснення класифікації.

3. Прогнози: цей етап стосується використання вже побудованого DT для прогнозів.

Рис. 3.2 зображує ADP більш детально. Вхідні дані переходять на етап попередньої обробки, а потім використовуються IWC для створення двох кластерів; один кластер об’єднує звичайні дані, а інший – аномальні. Кластеризація відбувається на основі подібності між атрибутами даних.

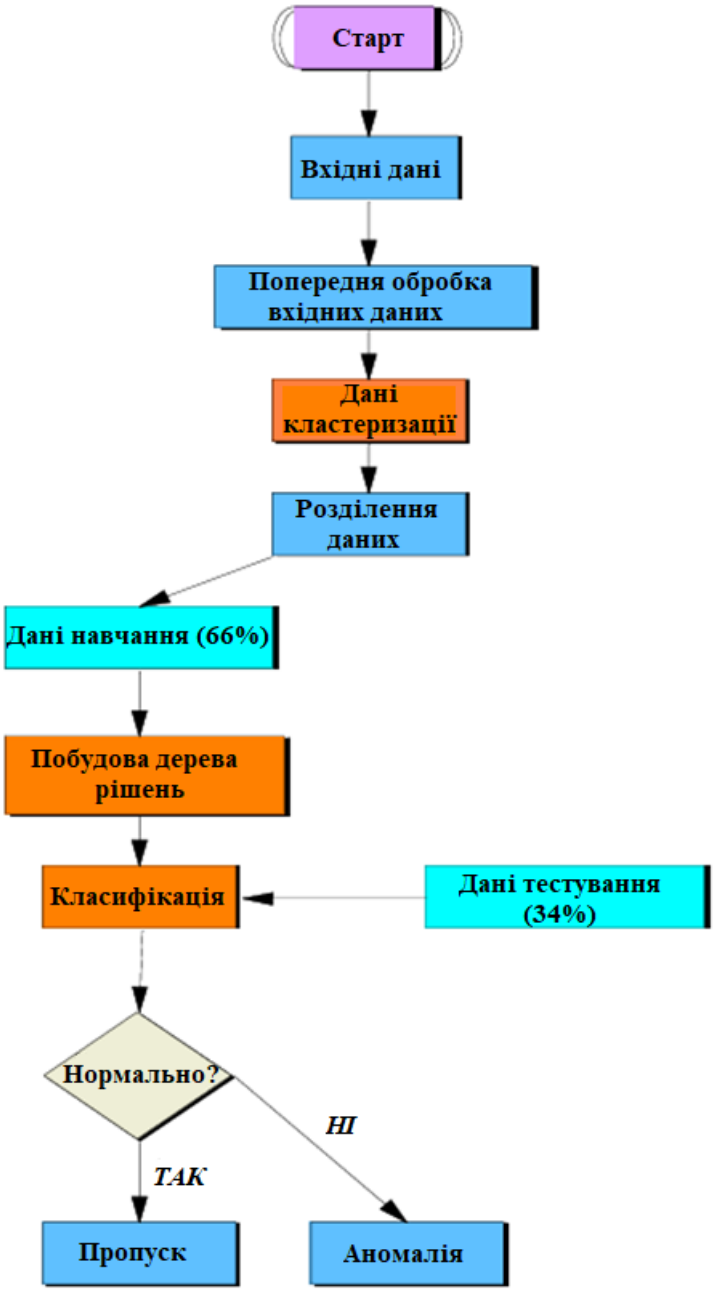

Рис 3.2 Блок-схема запропонованої моделі виявлення вторгнень.

Після кластеризації попередньо оброблені вхідні дані розбиваються на дві групи; одна група використовується для навчання та побудови дерева рішень, а інша група використовується для перевірки точності дерева рішень у розрізненні нормальних і ненормальних даних.

Вхідні дані. Коли справа доходить до вхідних даних для виявлення вторгнень в IoT, дуже важко отримати позначені дані. Дослідники вказує на цю проблему, з якою стикаються більшість незалежних дослідників у галузі виявлення вторгнень

в IoT. З іншого боку, дослідникам, які працюють над виявленням вторгнень, доступно безкоштовно багато наборів даних, не пов’язаних із IoT, наприклад, опубліковане Міністерством оборони; Набір даних KDD Cup. Набір даних складається з мільйонів записів, зібраних у військовому мережевому середовищі та призначених насамперед для досліджень виявлення мережевих вторгнень.

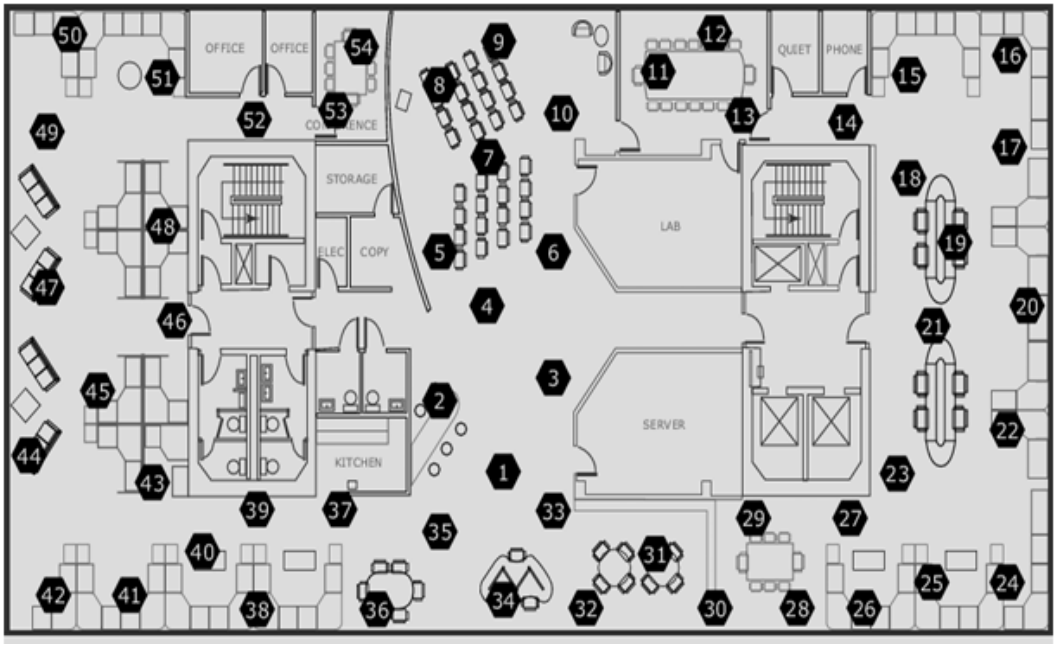

Рис. 3.3 Розподіл датчиків на приколаді приміщення офісу

На рис. 3.3 зображено розташування 54 датчиків у приміщенні. Кожен із цих датчиків може збирати дані про температуру, вологість, освітленість і напругу та передавати ці дані на шлюз кожні 31 секунду. Кожне зчитування розрізняється за датою, часом і ідентифікатором датчика відправника. Хоча цей набір даних є у вільному доступі, він не призначений перш за все для досліджень виявлення вторгнень, тому він не позначений даними. Тому він потребує обробки, щоб використовувати його для розробки запропонованої моделі виявлення аномалії в цьому дослідженні.

Попередня обробка даних. Обробка понад 2,3 мільйона показань 54 датчиків може бути дорогою з обчислювальної точки зору. Тому випадкова вибірка показань була взята з 2,3 мільйона показань і поміщена у файл .csv Microsoft Excel під назвою «inputData.csv». На додаток до цього непотрібні атрибути (тобто

стовпці), такі як дата та час зчитування, ідентифікатор датчика, світло та напруга, були даними, які були видалені, оскільки запропонована модель виявлення аномалій пов'язана з виявленням аномалій на прикладному рівні. Крім того, аномалії були додані до вхідних даних вручну (тобто 454 записи). Дослідники [21] використовували той самий спосіб для маніпулювання вхідними даними, додаючи вручну записи з аномаліями в наборі даних. Переваги такого підходу: (1) вхідні дані будуть якось позначені та відомі вчителю, (2) вхідний набір даних стане кваліфікованим набором даних для виявлення аномалій і відповідатиме меті дослідження. (3) подолання проблеми отримання міченого набору даних про вторгнення для IoT за дорогою ціною.

	Date	Time	Reading Sq.	Sensor ID	Temp.	Humidity	Light	Voltage
1	2024-03-31	03:38:15.757551	2	1	122.153	-3.91901	11.04	2.03397
2	2024-02-28	00:59:16.02785	3	1	19.9884	37.0933	45.08	2.69964
3	2024-02-28	01:03:16.33393	11	1	19.3024	38.4629	45.08	2.68742
4	2024-02-28	01:06:16.013453	17	1	19.1652	38.8039	45.08	2.68742
5	2024-02-28	01:06:46.778088	18	1	19.175	38.8379	45.08	2.69964
6	2024-02-28	01:08:45.992524	22	1	19.1456	38.9401	45.08	2.68742
7	2024-02-28	01:09:22.323858	23	1	19.1652	38.872	45.08	2.68742

Вхідні дані перед попередньою обробкою (необроблені дані)

	Temp.	Humidity
1	18.071	53.919
2	18.06	54.219
3	18.05	54.519
4	18.039	54.819
5	18.029	55.119
6	18.018	55.419
7	18.008	55.719

Вхідні дані після попередньої обробки

Рис.3.4 Вхідні дані до та після попередньої обробки

Отриманий вхідний набір даних складатиметься з двох атрибутів (тобто стовпців); температури та вологості та містить 7526 записів. Значення кожного атрибута в кожному записі є дійсним, як показано на рис.3.4.

3.2 Аналіз застосування кластеризації та класифікації

Відповідно аналізу, кластеризація або кластерний аналіз стосується поділу даних на групи таким чином, щоб елементи в кожній групі були схожі один на

одного, а не елементи в інших групах. Кожна група відома як кластер. Існує декілька типів підходів до кластеризації; кластеризація на основі центроїду, кластеризація на основі розподілу, кластеризація на основі щільності тощо. Кластеризація на основі центроїда зазвичай використовується в програмі виявлення вторгнень, оскільки вона представляє кожен кластер даних центральним вектором, який називається «центроїд». Центроїд не обов'язково є членом даних кластера. Це результат застосування алгоритму кластеризації до даних. K-means на сьогоднішній день є найпопулярнішим алгоритмом кластеризації. Враховуючи кількість кластерів, які потрібно створити, алгоритм може створити центроїди. Наприклад, якщо вхідні дані мають бути згруповані у дві групи – у цьому випадку k дорівнює 2, K-means випадковим чином вибере два записи з вхідних даних і почне отримувати інші записи в тому самому наборі даних і порівнювати їх з 2 випадково вибраними записами шляхом обчислення відстані між ними. Коли k-means обчислює відстань між записами, коротка відстань вказує на те, що отриманий запис, швидше за все, буде схожий на один із записів, вибраних випадковим чином на початку операції кластеризації. Один алгоритм завершує порівняння всіх записів; він створює k центроїдів, що представляють k кластерів. На цьому етапі кластери не маркуються. Тому вчитель повинен знайти спосіб позначити ці кластери. На рис.3.5 зображено кластерний набір даних у 3 групи за допомогою k-середніх.

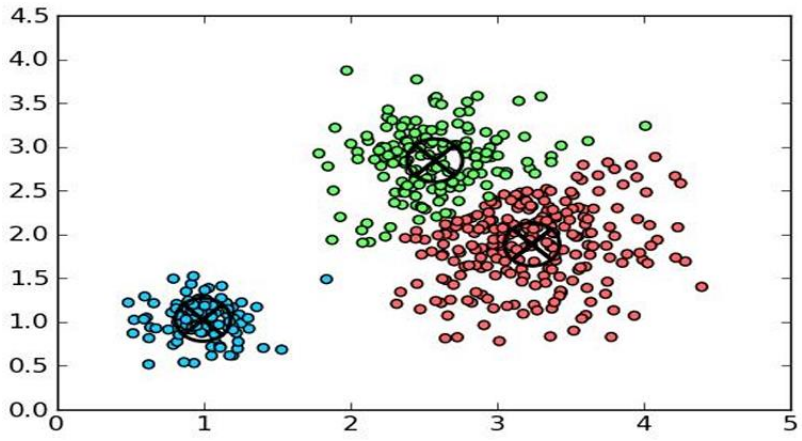

Рис.3.5 Приклад кластеризації K-середніх

Досліджники вказали на один із недоліків K-середніх. Оскільки вчитель, швидше за все, не знає, які записи чи k векторів найкраще представляють дані, і що kmeans випадковим чином вибирає ці записи, це означає, що центроїд, який створює алгоритм наприкінці кластеризації, може бути неточним. Тому вони запропонували алгоритм інверсної вагової кластеризації (IWC) для усунення цього критичного дефекту в K-середніх. IWC, по суті, побудований на алгоритмі k-середніх. Однак він покладається на виконання k-середніх багато разів, доки центроїди та кластери не стануть стабільними. Іншими словами, у той час як k-середні зупиняються після формулювання k-центроїдів і кластерів, IWC бере отримані центроїди та повторно запускає k-середні для тих самих даних, обчислюючи відстань між кожним записом і центроїдами. У K-середніх центроїди можна знайти за такою формулою:

$$m_k = \frac{\sum_n r_{kn} x_n}{\sum_{j,n} r_{jn}} \tag{3.1}$$

де

$$r_{kn} = \frac{\exp\left(-\beta d(x_n, m_k)\right)}{\sum_j \exp\left(-\beta d(x_n, m_j)\right)} \tag{3.2}$$

d(a,b) — евклідова відстань між a і b. IWC розширює K-means наступним чином:

$$J_I = \sum_{i=1}^{N} \sum_{k=1}^{K} \frac{1}{\left||x_i - m_k|\right|^P} \tag{3.3}$$

$$\frac{\partial J_I}{\partial m_k} = \sum_{i=1}^{N} P(x_i - m_k) \frac{1}{\left||x_i - m_k|\right|^{P+2}} \tag{3.4}$$

$$\frac{\partial J_I}{\partial m_k} = 0 \Rightarrow$$

$$m_k = \frac{\sum_{i=1}^{N} \frac{1}{\left||x_i - m_k|\right|^{P+2}} x_i}{\sum_{i=1}^{N} \frac{1}{\left||x_i - m_k|\right|^{P+2}}} = \frac{\sum_{i=1}^{N} b_{ik} x_i}{b_{ik}} \tag{3.5}$$

де

$$b_{ik} = \frac{1}{||x_i - m_k||^{P+2}} \quad (3.6)$$

Спосіб, за допомогою якого k-середні можна покращити, щоб стати менш чутливими до вибору початкового центроїда, можна виконати шляхом часткового виведення J_I відносно m_k, як показано в рівнянні (3.4). Це дасть центроїди, які найкраще представляють дані, як показано на рис. 3.6 .

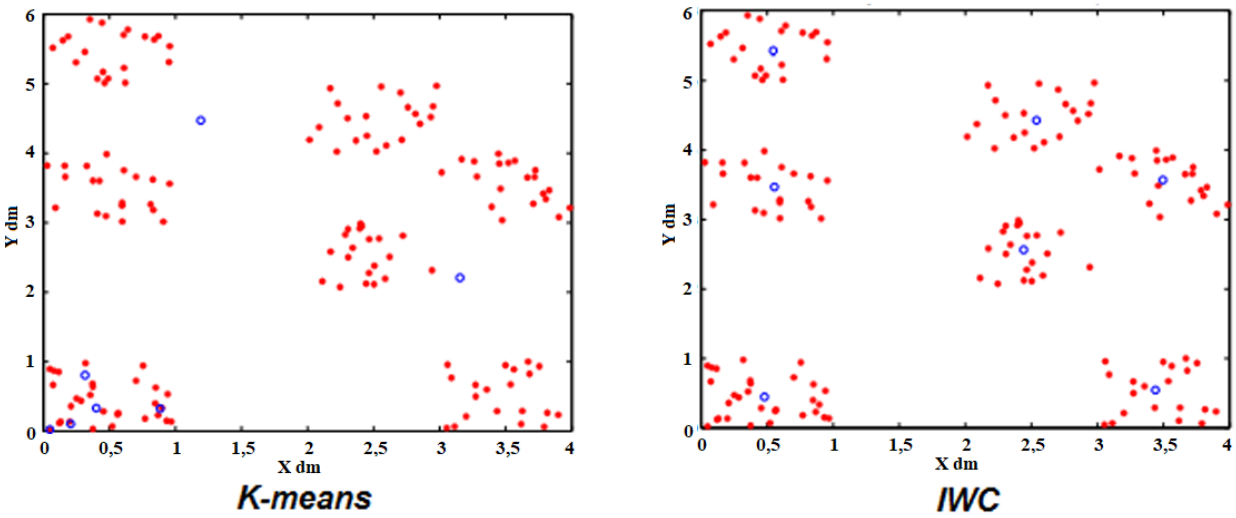

Рис.3.6 IWC проти K-середніх у даних кластеризації

Наведені вище рівняння для IWC, запропоновані Ashour і Fyfe (2007), можуть науково покращити кластеризацію k-середніх. Таким чином, ці рівняння будуть реалізовані та використані для кластеризації вхідних даних у запропонованій моделі виявлення аномалій для IoT, щоб створити центроїди, які найкращим чином представляють вхідні дані.

Класифікація. Алгоритми дерева рішень (DT) зазвичай використовуються в класифікації. DT можна сприймати як модель прогнозування, яка відображає спостереження щодо елемента до висновку в структурі дерева, що складається з вузлів і гілок, як показано на рис. 3.7. Кожен лист у дереві представляє клас. У випадку запропонованої моделі є два класи; нормальний і ненормальний. Ці класи виділяються в результаті кластеризації даних за допомогою IWC. Кожна гілка представляє сполучення, яке веде до мітки класу. C4.5 — це алгоритм, який можна

використовувати для створення DT. Його запропонував Росс Квінлан. Для того, щоб C4.5 міг побудувати DT, йому потрібно надати набір навчальних даних уже класифікованих зразків. Таким чином, вхідний набір даних, який використовується в кластеризації, можна розділити на дві частини:

1. Навчальний набір даних: містить 4990 записів, класифікованих як нормальні та ненормальні.

2. Тестовий набір даних: який містить 2536 записів без міток, щоб його можна було використовувати для тестування DT-C4.5.

Використовуючи навчальний набір даних, C4.5 починає будувати дерево. Вузол у дереві складається з фактичного значення температури та вологості. Базуючись на теорії нормалізації, C4.5 почне створювати інші листи в деревах, які можуть з'єднувати всі атрибути один з одним за допомогою нормальних і ненормальних класів для створення гілок за допомогою операторів if. Іншими словами, якщо значення даного атрибута дорівнює x, тоді клас буде нормальним або ненормальним на основі отримання інформації.

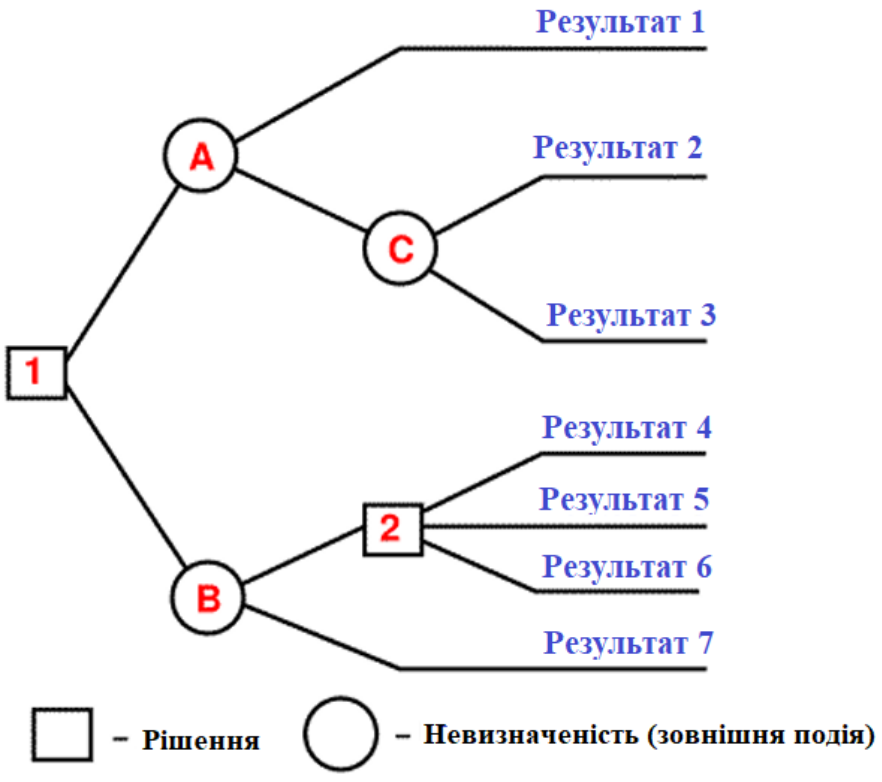

Рис.3.7 Дерево рішень

Псевдокод алгоритмів C4.5 показаний на рис. 3.8. Алгоритм приймає навчальний набір секретних даних. Спочатку перевіряється, чи задовольняються критерії завершення. Якщо так, алгоритм припиняє роботу. Якщо ні, алгоритм обчислює теоретико-інформаційні критерії для всіх атрибутів. Потім він вибирає найкращий атрибут, використовуючи інформаційно-теоретичні критерії. Потім він створює вузол рішення на основі найкращого атрибута, який буде першим вузлом у дереві. З першого вузла алгоритм починає розділяти навчальний набір даних на основі першого вузла на наступні вузли та рекурсивно обробляє кожен атрибут від найнижчого рівня в дереві до верхнього вузла, продовжуючи оновлювати всі вузли, доки не буде виконано критерії завершення.

Algorithm 1.1 C4.5(D)

Input: an attribute-valued dataset D

```
1: Tree = {}
2: if D is "pure" OR other stopping criteria met then
3:    terminate
4: end if
5: for all attribute a ∈ D do
6:    Compute information-theoretic criteria if we split on a
7: end for
8: a_best = Best attribute according to above computed criteria
9: Tree = Create a decision node that tests a_best in the root
10: D_v = Induced sub-datasets from D based on a_best
11: for all D_v do
12:    Tree_v = C4.5(D_v)
13:    Attach Tree_v to the corresponding branch of Tree
14: end for
15: return Tree
```

Рис.3.8 Алгоритми C4.5

Метод оцінки. Матриця плутанини — це техніка, яка зазвичай використовується для опису та характеристики ефективності моделі класифікації в системі виявлення вторгнень. Матриця проста, але містить заплутані елементи; істинно позитивний (TP), істинно негативний (TN), хибнопозитивний (FP) і хибнонегативний (FN). Це розмірна матриця 2x2, яка представляє співвідношення між фактичним значенням і прогнозованим значенням, як показано на рис. 3.9. Якщо класифікатор передбачає, що в даних є проблеми чи аномалії, і дані насправді є ненормальними, тоді спроба вважається істинно позитивною (TP).

У випадку запропонованої моделі виявлення аномалій, якщо модель передбачає, що певні тестові дані або запис є аномалією, а дані або запис насправді є аномалією, тоді це можна використовувати як індикатор швидкості TP запропонованої системи

Рис.3.9 Матриця плутанини

Якщо класифікатор передбачає, що в даних немає проблем і ці дані насправді нормальні, тоді спроба вважається істинно негативною (TN). У випадку запропонованої моделі виявлення аномалій, якщо модель передбачає, що певні тестові дані чи запис є нормальними, а дані чи запис насправді є нормальними, тоді це можна використовувати як показник швидкості TN запропонованої системи. Якщо класифікатор передбачає наявність проблеми чи аномалії в даних, хоча дані насправді нормальні й не містять проблеми чи аномалії, тоді спроба вважається хибнопозитивною (FP). У випадку запропонованої моделі виявлення аномалій, якщо модель передбачає, що певні тестові дані чи запис мають проблему чи аномалію, хоча їх немає, тоді це можна використовувати як показник швидкості FP запропонованої системи. Більшість моделей виявлення аномалій сподіваються знизити FP до найнижчого рівня. Нарешті, якщо класифікатор передбачає, що в даних немає проблеми чи аномалії, тоді як у даних насправді є проблема чи аномалія, тоді спроба вважається хибнонегативною (FN). У випадку запропонованої моделі виявлення аномалій, якщо модель передбачає, що певні

тестові дані чи запис не мають проблеми чи аномалії, хоча й мають, це можна використовувати як індикатор швидкості FN запропонованої системи. У моделях виявлення аномалій цей показник вказує на рівень нездатності моделі забезпечити заплановану функцію – виявлення аномалій. Маючи це на увазі, загальну точність системи – як часто класифікатор правильний – можна обчислити за такою формулою:

$$\frac{TP + TN}{x}$$

Де x — загальна кількість записів, наданих як вхідні дані для моделі класифікації. У випадку цього дослідження x дорівнює кількості записів у даних тестування; 2,536. Рівень неправильної класифікації моделі – як часто класифікатор помиляється, можна обчислити за такою формулою:

$$\frac{FP + FN}{x}$$

Швидкість TP (він же Чутливість або Відкликання) моделі можна обчислити за такою формулою:

$$\frac{TP}{number\ of\ actual\ abnormal\ records\ in\ x}$$

Швидкість FP моделі можна обчислити за такою формулою:

$$\frac{FP}{number\ of\ actual\ normal\ records\ in\ x}$$

Специфіку моделі або швидкість TN можна розрахувати за такою формулою:

$$\frac{TN}{number\ of\ actual\ normal\ records\ in\ x}$$

Точність моделі можна обчислити за такою формулою:

$$\frac{TP}{number\ of\ prediced\ abnormal\ records\ in\ x}$$

Нарешті, поширеність моделі можна обчислити за такою формулою:

$$\frac{actual\ number\ of\ abnormal\ records\ in\ x}{x}$$

Ці показники будуть використані для оцінки запропонованої моделі виявлення аномалій і порівняння результатів з іншими пов'язаними роботами.

3.3 Впровадження моделі для виявлення аномалій

Реалізація. Запропонована модель для виявлення аномалій в IoT складається з шести файлів. Файли описані в табл.3.1. та додаються до додатку.

Таблиця 3.1

Файли, які створюють запропоновану модель для виявлення аномалій в IoT

Ім'я файлу	опис
IoT_Clustering.m	Цей MATLAB-файл реалізує етап кластеризації в ADP. Він приймає попередньо оброблений вхідний файл у форматі .csv, перетворює його в матрицю, викликає IWK.m для кластеризації даних на дві групи, генерує центроїди і створює кластеризовані дані у вигляді файлу..csv. .csv – це файл Microsoft Excel, який може обробити MATLAB.
IWK.m	Цей файл MATLAB представляє реалізацію IWC, запропоновану Ashour і Fyfe (2007). Він приймає вхідні дані у формі матриці, виводить центроїди на основі 47 кількості кластерів K і повертає отримані центроїди у формі матриці.

Продовження таблиці 3.1

Файли, які створюють запропоновану модель для виявлення аномалій в IoT

Ім'я файлу	опис
IoT_Classification.m	Цей MATLAB-файл реалізує етап класифікації в ADP. Він отримує кластеризовані дані у вигляді матриці та розподіляє їх на дві групи: навчальну та тестову. Навчальна група застосовується для створення дерева рішень.
C4_5.m	Цей MATLAB-файл реалізує алгоритм C4.5 для побудови дерева рішень. Він приймає матрицю з навчальними даними, матрицю ідентифікаторів кластерів, що визначають кожен запис у навчальній матриці, а також кількість початкових вузлів дерева. За допомогою файлу make_tree.m він створює дерево рішень, застосовуючи інформаційно-теоретичні критерії. Файл C4_5.m генерує дерево рішень.
make_tree.m	Цей MATLAB-файл створює дерево.
use_tree.m	Цей MATLAB-файл використовує дерево рішень, створене за допомогою файлів C4_5.m і make_tree.m, та застосовує його до тестових даних. Він також порівнює класифікаційні результати вхідних даних з ідентифікаторами кластерів для кожного запису в тестовому наборі, щоб отримати кінцеві результати. Результати представлені у вигляді істинно позитивних, істинно негативних, хибнопозитивних і хибнонегативних значень.

MATLAB використовується для впровадження та тестування запропонованої моделі для виявлення аномалій в IoT. Він відносно простий і легкий у використанні порівняно з іншими мовами програмування високого рівня, такими як Java або C. Матриця є основним елементом у MATLAB. Навіть просте ціле число розглядається як матриця з одним рядком і одним стовпцем. З огляду на типи вхідних даних, що використовуються в задачах виявлення аномалій, дані зазвичай подаються у вигляді матриць. MATLAB також характеризується високою швидкістю обробки та здатністю працювати з великими обсягами даних. Він автоматично розподіляє обчислення між процесорними ядрами без необхідності

оптимізації або використання розподіленої обробки, як це відбувається в інших мовах програмування.

Попередній друк вхідних даних. Набір даних Intel Lab IoT, описаний у попередньому розділі, складається з 2,3 мільйона показань 54 датчиків, розподілених у кількох місцях офісу. Датчики збирають 4 типи сенсорних даних; температура, вологість, світло та напруга. Після завантаження набір даних надходить у текстовому форматі. Його можна відкрити за допомогою Microsoft Excel. Дані виглядатимуть так, як показано на рис.3.10. Перші 4 стовпці в цих даних представляють дату та час збору даних, читання та ідентифікатор вузла відповідно. Ці стовпці або атрибути не потрібні. Тому ці атрибути будуть виявлені. Останні 4 стовпці або атрибути представляють температуру, вологість, світло та напругу відповідно. Перші два атрибути необхідні в запропонованій моделі для виявлення аномалій в IoT. Інші два атрибути можуть бути оброблені в рамках майбутньої роботи.

2024-03-31	03:38:15.757551	2	1	122.153	-3.91901	11.04	2.03397
2024-02-28	00:59:16.02785	3	1	19.9884	37.0933	45.08	2.69964
2024-02-28	01:03:16.33393	11	1	19.3024	38.4629	45.08	2.68742
2024-02-28	01:06:46.778088	18	1	19.1652	38.8039	45.08	2.68742
2024-02-28	01:08:45.992524	22	1	19.175	38.8379	45.08	2.69964

Рис.3.10 Зразок необроблених даних, отриманих від Intel Lab

Після видалення зайвих стовпців і даних з набору даних Intel Lab, новий набір матиме формат, представлений на рис. 3.11, з двома стовпцями (атрибутами), що відображають температуру та вологість. З 2,3 мільйона вимірювань для створення вхідного набору було вибрано випадковим чином 7072 показання. Додатково до цього набору було включено ще 454 показання з аномаліями. Отже, в результаті вхідний набір міститиме 7526 показань.

122.153	-3.91901
19.9884	37.0933
19.3024	38.4629
19.1652	38.8039
19.175	38.8379

Рис.3.11 Зразок вхідних даних після попередньої обробки

Реалізація алгоритму IWC Алгоритм Invers Weight Clustering (IWC) імплантовано у файл IWC.m. Метод реалізовано у вигляді функції MATLAB, яка приймає два аргументи:

1. Матрицю вхідних даних, що підлягають кластеризації.
2. Кількість кластерів, які алгоритм має сформувати.

Функція повертає матрицю кластеризованих даних. Інтерфейс функції задається наступним рядком коду (рядок 4: IWC.m).

функція [m] = IWK(x,K)

Модель створена для реалізації функціоналу IWC з використанням файлу IoT_Clustering.m. У цьому файлі реалізовано читання вхідних даних у форматі .csv та їхнє перетворення на матрицю розміром 7526x2. Для цього використовується функція csvread(), яка є стандартним інструментом MATLAB для автоматичного зчитування даних із файлів .csv і перетворення їх у матричний формат. У наведених нижче рядках коду продемонстровано, як вхідний файл зчитується і конвертується у матрицю (рядки 1 та 3: IoT_Clustering.m).

ім'я файлу = 'inputData.csv'; M = csvread(назва файлу);

Після обробки файлу Excel в MATLAB, вхідні дані будуть збережені в матриці в системі пам'яті та підготовлені для кластеризації. Для зручності матриця вхідних даних буде представлена у вигляді, аналогічним даним на рис. 3.12.

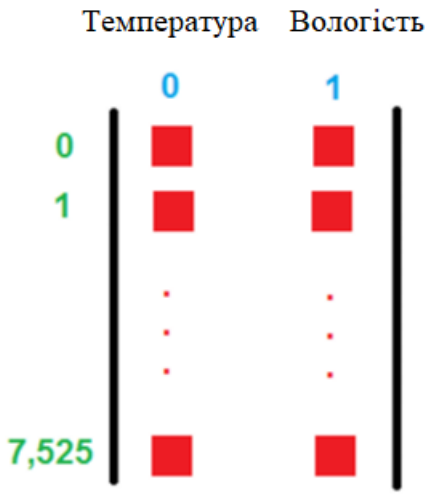

Рис.3.12 Матриця вхідних даних перед кластеризацією

Матриця даних, зображена на рис.3.12, буде передана функції IWC разом із кількома кластерами, які потрібно створити на базі слідуючого рядка коду (рядок 8: IoT_Clustering.m):

$$C = IWK(M,K)$$

Де M є матрицею вхідної інформації, а K дорівнює 2 кількістю кластерів, що повинен створити алгоритм IWC. IWC обирає два випадкові рядки з вхідної матриці для порівняння з іншими рядками, використовуючи обчислення евклідової відстані. Це дозволяє створити кластери та визначити центроїди. Для отримання оптимального центроїда, що найкраще описує два кластери, IWC виконує цю процедуру 10 разів із нормалізацією даних (рядки 20-85: IWC.m). У результаті, IWC генерує матрицю, яка містить центроїд і два рядки, що описують значення відповідних похідних центроїдів.

Ця матриця повертається функцією IoT_Clustering.m, яка застосовує аналогічний метод нормалізації, обчислюючи відстань між кожним рядком вхідної матриці та центроїдами. На основі цих відстаней, IoT_Clustering.m відтворює матрицю вхідних даних і призначає кожному рядку мітку кластеру: «1» або «2». Мітка «1» вказує на належність запису до першого кластеру, а «2» — до другого.

Кінцевий результат — кластеризована матриця даних, подібна до тієї, що наведена на рисунку 3.13.

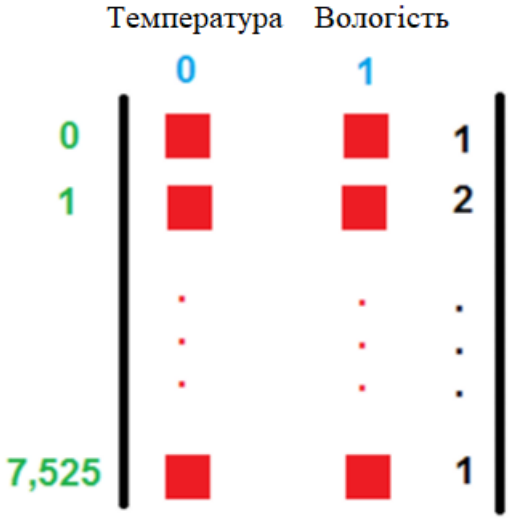

Рис.3.13 Матриця вхідних даних після кластеризації

IoT_Clustering.m перетворює кластерну матрицю вхідних даних у файл Excel із застосуванням функції MATLAB csvwrite(). Це робиться в наступному рядку коду (рядок 27: IoT_Clustering.m):

csvwrite(reslutfile,R)

Resultfile — це змінна, яка зберігає назву Excel-файлу, що містить кластерні вхідні дані, а R — це назва матриці, де розміщено ці кластеризовані дані. Код призначений для збереження даних у файлі Excel з назвою «clusteredData.csv». Приклад вмісту цього файлу наведено на ілюстрації. Експорт кластерних вхідних даних у Excel-файл здійснюється для подальшого використання їх на наступному етапі — побудови дерева рішень (DT) та проведення тестування.

38.453	56.019	2
39.353	56.319	2
40.253	56.619	2
38.453	39.144	1
39.353	39.144	1
40.253	39.042	1
32.153	39.144	1

Рис.3.14 Приклад вхідної інформації після процесу кластеризації

Застосування класифікатора – C4.5. C4.5 — це алгоритм побудови дерева рішень, який використовується для класифікації ймовірностей на основі наявних даних. У файлі IoT_Classification.m реалізовано процес класифікації з використанням алгоритму C4.5. Перед створенням дерева рішень дані розділяють на дві групи: навчальну та тестову. Навчальна група містить 4990 записів (66% кластеризованих вхідних даних), а тестова — 2536 записів (34% кластеризованих вхідних даних). Файл IoT_Classification.m читає кластерні дані, збережені у форматі Excel, за допомогою функції MATLAB csvread(), і зберігає їх у вигляді матриці. Цей процес відображено у наступних рядках коду (рядки 2-3: IoT_Classification.m).

```
Дані = 'clusteredData.csv';
CData = csvread(Дані).
```

Матриця, яку створить IoT_Classification.m, точно схожа на ту, що зображена на рис.3.13 Однак IoT_Classification.m розділяє цю матрицю на дві матриці; матриця даних навчання та матриця даних тестування. Ці дві матриці не містять ідентифікатора кластера кожного рядка. Те, як виглядатимуть ці матриці, зображено на рис.3.15. Крім того, вони створюються за допомогою таких рядків коду (рядок 11 і 14: IoT_Classification.m):

```
train_data = CData(1:4990,1:2)'
test_data = CData(4991:7526,1:2)'
```

IoT_Classification.m розділяє кластерну матрицю введення даних на дві додаткові матриці; навчальна матриця цільових даних і тестова матриця цільових даних. Ці дві матриці складаються з одного атрибута, який є ідентифікатором кластера кожного рядка. Те, як виглядатимуть ці матриці, зображено на рис.3.16 Крім того, вони створюються за допомогою таких рядків коду (рядки 12 і 15: IoT_Classification.m):

```
train_targets = CData(1:4990,3)'
test_targets = CData(4991:7526,3)'
```

Знак ««» у кінці кожної матриці вказує на те, що матриця в зворотному порядку, оскільки C4.5 вимагає, щоб вхідні матриці надавалися у зворотному порядку під час побудови дерева. Дані тренування та матриці цільових даних надсилаються до C4_5.m разом із відсотком неправильно призначених вибірок у вузлі. Це можна побачити в IoT_Classification.m у рядку 17 коду:

```
[дерево, discrete_dim] = C4_5 (train_data, train_targets, 1)
```

Для кожного рядка в матриці даних поїзда C4_5.m обробляє кожен запис, доки не знайде таку кількість вузлів, яку міститиме дерево. Це вважається критерієм припинення. Після отримання цього числа дані тренування, цільові дані тренування та максимальна кількість вузлів у дереві використовуються для побудови дерева рішень за допомогою make tree.m, який викликається з C4_5.m у рядку 21: tree = make_tree(train_data, train_targets, inc_node, discrete_dim, max(discrete_dim), 0);

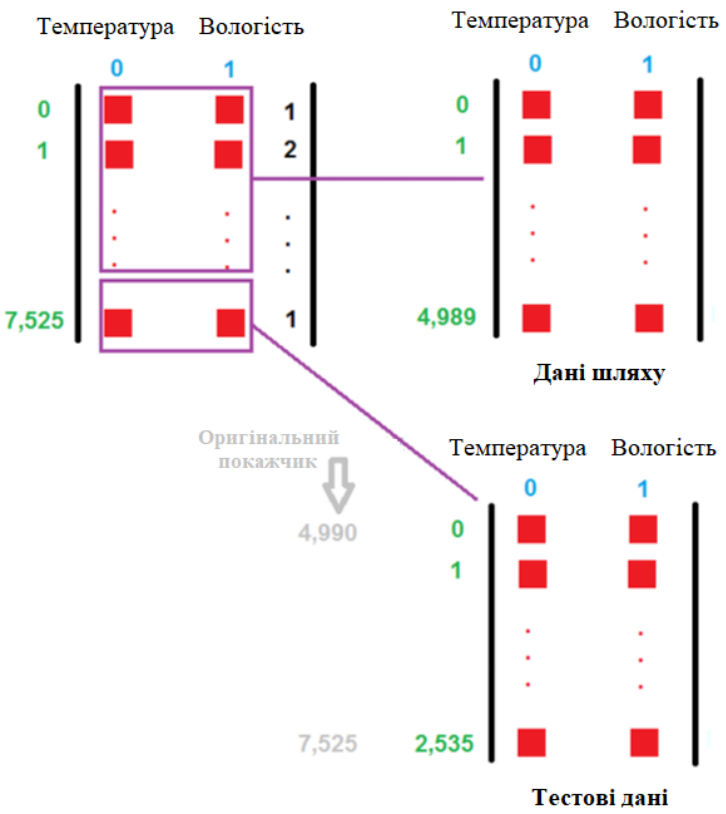

Рис.3.15 Розділення вхідної матриці в тренувальних і тестових даних

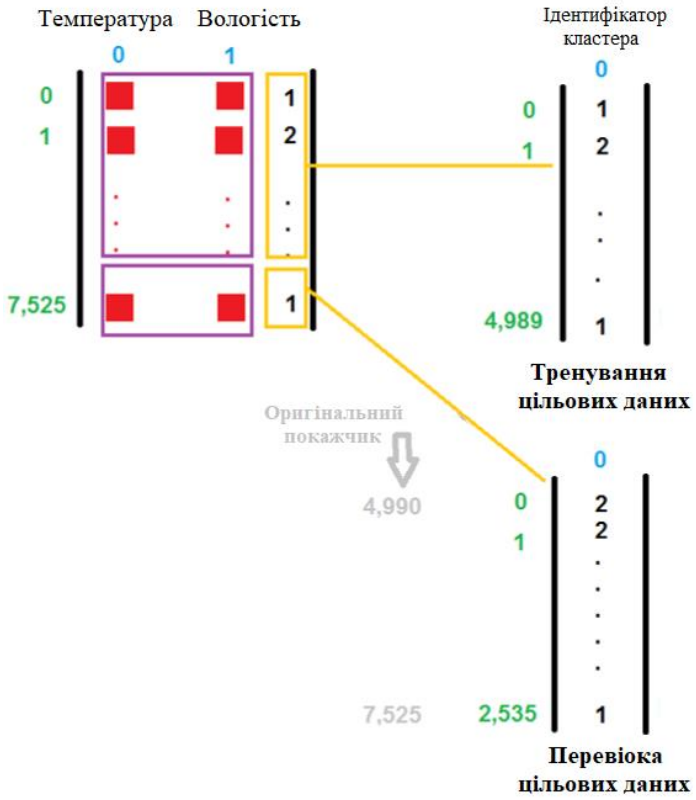

Рис.3.16 Розділення вхідної матриці в цільових даних поїзда та тестових цільових даних

3.4 Результати застосування моделі для виявлення аномалій

Після запуску моделі за допомогою MATLAB отримані результати кластеризації та класифікації даних. На рис.3.17 і рис.3.18 показано вхідні дані до і після кластеризації за допомогою алгоритму IWC.

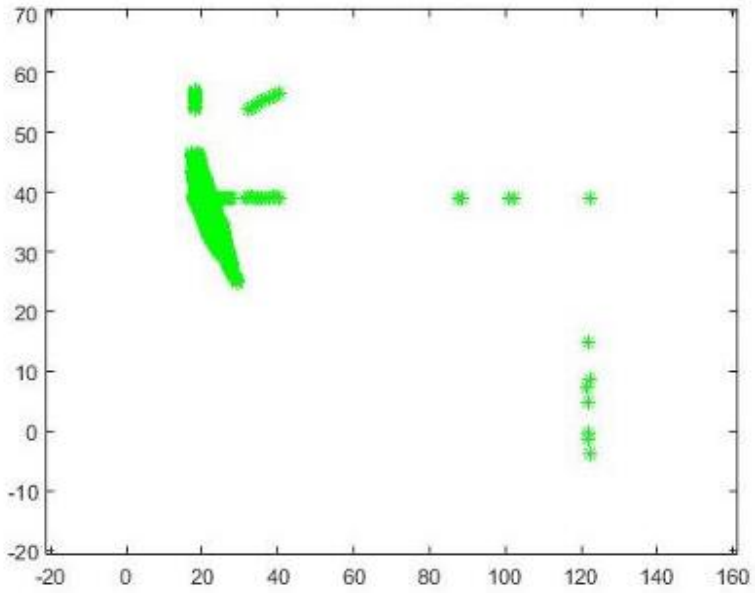

Рис.3.17 Вхідні дані перед кластеризацією.

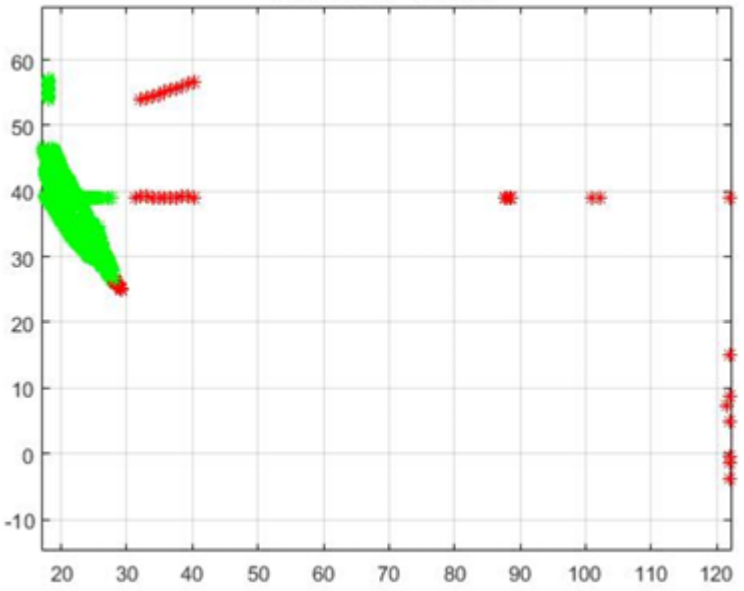

Рис.3.18 Вхідні дані після кластеризації

На рис.3.17 видно, що всі дані належать одному класу або кластеру. Після кластеризації вхідних даних за допомогою алгоритму IWC дані були розділені на два класи, один показаний червоним кольором, що вказує на те, що ці дані належать до класу 1, а інші дані належать до класу 2 і показані зеленим кольором на рис.3.18. Після запуску класифікатора на тестових даних, які складаються з 2536 рядків, модель повідомила параметри оцінки матриці плутанини; TP, TN, FP і FN.

Табл.3.2 показує результати запропонованої моделі виявлення аномалій на матриці плутанини.

Таблиця 3.2

Результати запропонованої моделі виявлення аномалій на матриці помилок

X= 2536	Прогнозовані аномалії	Прогнозована нормальність
Фактичні аномалії (182)	TP = 179	FN = 3
Фактична норма (2354)	FP = 51	TN = 2303

Оцінка результатів. У цьому розділі представлено оцінку запропонованої моделі виявлення аномалій в IoT. Він містить розраховані показники ефективності запропонованої моделі у світлі результатів, наведених у попередньому розділі. Показники продуктивності – це показники, які можна обчислити, використовуючи кількість правильних і неправильних передбачень, зроблених моделлю, і включають загальну точність, частоту неправильної класифікації, чутливість, специфічність, точність і поширеність. Аналізуючи модель для виявлення аномалій в IoT із застосуванням алгоритму (IWC) і алгоритму дерева рішень C4.5 виконана перевірка із застосуванням позначеного тестового набору інформації, яка налічує 2536 записів (тобто 34% вхідного набору даних). Кожен із таких запис у тестовому наборі даних містить три значення; температура, вологість та ідентифікатора кластерів. Температурні показники та показники вологості, які спершу були передані з датчиків IoT у офісі. Ідентифікатор кластера – це ціле число, що може набувати значення «1» чи «2». Дані значення виконують представлення кластеру,

до якого відноситься кожне з показань (а саме температурні дані і дані вологіості). Показники ID кластера що отримані по результатам застосування алгоритму IWC проти вхідних наборів інформації, що складається з кількості 7526 виконаних записів. Дерево рішень C4.5 було навчено за допомогою позначеного навчального набору даних, і є складанням 4990 записів. Далі його було запущено з тестовим набором даних, але без надання ідентифікатора кластера, який показує, до якого кластера належить кожен запис у тестовому наборі даних. Іншими словами, C4.5 було запущено проти тестового набору інформації які не мають міток. Оскільки кількість записів у немаркованому наборі даних тестування становить 2536, модель створила 2536 прогнозів у формі «1» або «2». Отримані прогнози порівнювали з ідентифікатором кластера кожного запису в позначеному наборі даних тестування, і результати показали:

- загальна кількість передбачених аномалій, які насправді були аномаліями, становила 179. Таким чином, справжнє позитивне (TP) значення моделі дорівнює:

$$TP = 179$$

- загальна кількість передбачених нормальних показників, які фактично були нормальними показниками, становила 2303. Отже, справжнє негативне (TN) значення моделі дорівнює:

$$TN = 2303$$

- загальна кількість передбачених аномалій, які фактично були нормальними, становила 51. Отже, хибнопозитивне (FP) значення моделі становить:

$$FP = 51$$

- загальна кількість передбачених нормальних показників, які насправді були аномаліями, становила 3. Таким чином, помилково негативне (FN) значення моделі дорівнює:

$$FN = 3$$

Для того, щоб мати можливість обчислити показники ефективності запропонованої моделі, такі як загальна точність, частота неправильної класифікації, чутливість, специфічність, точність і поширеність, потрібні такі фактичні значення: Нехай x відноситься до загальної кількості записів у тестовому наборі даних , тоді

$$x = 2536$$

Фактична кількість норм у наборі даних тестування становить 2354. Нехай A_{Normal} стосується фактичної кількості норм у наборі даних тестування, тоді

$$A_{Normal} = 2{,}354$$

Фактична кількість аномалій у наборі даних тестування становить 182. Нехай $A_{Anomalies}$ стосується фактичної кількості аномалій у наборі даних тестування, тоді

$$A_{Anomalies} = 182$$

Таким чином, загальна точність запропонованої моделі, яка дає правильні прогнози, становить 0,97.

$$\frac{\mathrm{TP+TN}}{\mathrm{x}} = \frac{179+2{,}303}{2{,}536} = 0{,}97$$

Коефіцієнт помилкової класифікації запропонованої моделі для створення неправильних прогнозів становить 0,021.

$$\frac{FP+FN}{x} = \frac{51+3}{2,536} = 0,021$$

Чутливість або відкликання запропонованої моделі при виявленні аномалій становить 0,98.

$$\frac{TP}{A_{Anomalies}} = \frac{179}{182} = 0,983$$

Коефіцієнт помилкових позитивних результатів запропонованої моделі становить 0,022.

$$\frac{FP}{A_{Normalities}} = \frac{51}{2,354} = 0,021$$

Специфічність запропонованої моделі в точному визначенні нормальних показань становить 0,97.

$$\frac{TN}{A_{Normalities}} = \frac{2303}{2354} = 0,97$$

Точність запропонованої моделі попередження при виявленні аномалій становить 0,78.

$$\frac{TP}{TP+FP} = \frac{179}{179+51} = 0,78$$

Поширеність запропонованої моделі становить 0,072.

$$\frac{A_{Normalities}}{x} = \frac{182}{2,536} = 0,072$$

ВИСНОВКИ

Найкращі програми Інтернету речей і найпопулярніші приклади використання IoT у різних галузях доводять, що IoT має руйнівний потенціал. Однією з найбільших переваг Інтернету речей є можливість змусити пристрої спілкуватися один з одним. Нові програми в ландшафті IoT сформують нову екосистему протоколів, пристроїв і мереж IoT.

Тому розумно очікувати можливості нових бізнес-моделей і кар'єрних можливостей у сфері IoT. Обізнаність щодо популярних програм Інтернету речей пропонує більше, ніж просто знайомство з IoT. Можливо дізнатися про різноманітність варіантів використання IoT та масштаби інновацій у впровадженні IoT у різних галузях.

Сектор IoT переживає трансформаційні зміни. Зміщення гіпермасштабувальників у бік периферійних і контейнерних стратегій, інтеграція штучного інтелекту в промислову автоматизацію, поява генеративного штучного інтелекту у виробництві та зростання рішень для обробки даних представляють лише деякі з динамічних подій, які переосмислюють корпоративну екосистему IoT. Ці тенденції, поряд з іншими ринковими даними, не лише відображають поточний стан ринку, але й дають уявлення про майбутнє зростання та інновації.

Орієнтуючись у цьому ландшафті, для компаній вкрай важливо залишатися поінформованими та адаптуватися. З прогнозованим CAGR у 17% до 2030 року потенціал для зростання та трансформації в секторі Інтернету речей є величезним.

Обґрунтування дослідження вторгнень і аномалій у системах IoT пояснюється збільшенням кількості атак у системах IoT. Представлено вичерпний огляд останніх робіт щодо схем виявлення аномалій на основі машинного та глибокого навчання для мереж IoT. Проаналізовано різні методи, які використовуються для досягнення основних результатів, і показано, як виявлення вторгнень для Інтернету речей має свої особливі проблеми через різнорідність пристроїв, обмежені обчислювальні можливості та величезну кількість підключених пристроїв, що ускладнює вбудовування системи виявлення вторгнень

складніше. Системи виявлення вторгнень (IDS) використовуються для захисту даних, якими обмінюються кінцеві точки та процеси в інформаційній системі, від несанкціонованого доступу та модифікації.

У третьому розділі виконано опис методології, спосіб, за допомогою якого модель виявлення аномалій в IoT використовує комбінацію двох алгоритмів, а саме алгоритму інверсної вагової кластеризації (IWC) і дерева рішень (DT). Він містить опис даних, які використовуютьсяя при реалізації та тестуванні моделі, а також опис того, як IWC і DT працюють у кластеризації та класифікації даних. Він також описує критерії оцінки, критерії, за якими можна виміряти точність моделі у виявленні аномалій в IoT.

У практичній частині, кваліфікаційної роботи, писано впровадження досліджуваної моделі для виявлення аномалій в IoT. Модель реалізовано та протестовано за допомогою MATLAB. Реалізація моделі складається з трьох етапів: попередня обробка вхідних даних, кластеризація, класифікація та тестування.

Щоб побудувати запропоновану модель виявлення аномалій і досягти другої мети цього дослідження, було отримано немаркований набір даних для IoT. Дані були попередньо оброблені, щоб містити лише два атрибути; температура і вологість. Кількість записів у наборі даних також зменшено до 7526 записів. Набір даних було згруповано за допомогою IWC у дві групи; кожна група була ідентифікована ідентифікатором зі значенням «1» або «2». Потім набір даних було розділено на два набори даних; навчальний набір даних, який містить позначені записи та становить 66% набору даних, і тестові дані, які містять немарковані записи та 62 представляють 34% набору даних. Навчальний набір даних використовувався C4.5 для побудови дерева рішень, яке потім застосовувалося до даних тестування. Результати показують, що загальна точність пропонованої системи у передбаченні аномалій і норм в IoT становить 97%, а рівень помилкової класифікації становить 2,1%. Чутливість системи при виявленні аномалій перевищує 98%, а частота помилкових позитивних результатів становить 2,1%.

ПЕРЕЛІК ПОСИЛАНЬ

1. Malhotra, P.; Singh, Y.; Anand, P.; Bangotra, D.K.; Singh, P. K.; Hong, W.C. Internet of things: Evolution, concerns and security challenges. Sensors 2021, pp. 78-80.

2. Berger, C.; Eichhammer, P.; Reiser, H.P.; Domaschka, J.; Hauck, F.J.; Habiger, G. A Survey on Resilience in the IoT: Taxonomy, Classification, and Discussion of Resilience Mechanisms. ACM Comput. Surv. 2022, pp.54-56.

3. Yousef, Z.; Barhoumi, H.G.W. Four-layer Architecture for IoT Security in Fog Network. In Proceedings of the 2023 10th International Conference on Internet of Things: Systems, Management and Security. October 2023; pp. 59–65.

4. Cviti, I.; Perakovi, D.; Periša, M.; Stojanovi, M.D. Novel classification of IoT devices based on traffic flow features. J. Organ. End User Comput. 2021, pp. 1–20

5. Gupta, A.; Fernando, X.; Das, O. Reliability and Availability Modeling Techniques in 6G IoT Networks: A Taxonomy and Survey. In Proceedings of the 2021 International Wireless Communications and Mobile Computing.2021 pp. 586–591

6. Goudarzi, M.; Palaniswami, M.; Buyya, R. Scheduling IoT Applications in Edge and Fog Computing Environments: A Taxonomy and Future Directions. ACM Comput. Surv. 2022, pp. 55.

7. Ahmed, H.I.; Nasr, A.A.; Abdel-Mageid, S.; Aslan, H.K. A survey of IoT security threats and defenses. Int. J. Adv. Comput. Res. 2019, pp. 325–350

8. Khanam, S.; Ahmedy, I.B.; Idris, M.Y.I.; Jaward, M.H.; Sabri, A.Q.B.M. A Survey of Security Challenges, Attacks Taxonomy and Advanced Countermeasures in the Internet of Things. 2020, pp. 124-127.

9. Victor, P.; Lashkari, A.H.; Lu, R.; Sasi, T.; Xiong, P.; Iqbal, S. IoT malware: An attribute-based taxonomy, detection mechanisms, and challenges. Peer Peer Netw. Appl. 2023, pp. 138–143.

10. Cook, A.A.; Misirli, G.; Fan, Z. Anomaly Detection for IoT Time-Series Data: A Survey. IEEE Internet Things J. 2020, pp. 648–649.

11. Aggarwal, C. C. & Reddy, C. K. Data Clustering: Algorithms and Applications. Chapman and Hall/CRC. 2023, pp. 11-17.

12. Alexander, D. & Finch, A. Information Security Management Principles (2nd Eds.). BCS, The Chartered Institute for IT. 2019, pp. 44-45.

13. Ashour, W. & Fyfe, C. Inverse Weighted Clustering Algorithm. The University of Paisley. 2017, pp. 34-36.

14. Attaway, S. Matlab: A Practical Introduction to Programming and Problem Solving (3rd Eds.). 2018. pp. 110.

15. Butterworth-Heinemann. Basu, S. K. Design Methods and Analysis of Algorithms (2nd ed.).2016, pp.20-22.

16. PHI Learning. Bell, J. Machine Learning: Hands-On for Developers and Technical Professionals. John Wiley & Sons. 2020, pp.54-60.

17. Buczak, A. L. & Guven, E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. In IEEE Communications Surveys & Tutorials. 2018, pp. 1153-1176.

18. Butun, I., Kantarci, B., & Erol-Kantarci, M. Anomaly detection and privacy preservation in cloud-centric Internet of Things. In IEEE International conference on Communication Workshop (ICCW), London, 8-12 June 2019, pp. 2610–2615.

19. Flach, P. Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press. 2020, pp.77-79.

20. Fortino, G. & Trunfio, P. (Eds.). Internet of Things Based on Smart Objects: Technology, Middleware and Applications. Springer. 2021, pp.121-124.

21. Fu, R., Zheng, K., Zhang, D., & Yang, Y. An Intrusion Detection Scheme Based on Anomaly Mining in Internet of Things. Nov. 2019, pp. 315-320.

22. Haq, N., Onik, A., Hridoy, A., Rafni, M., Shah, F., & Farid, D. Application of Machine Learning Approaches in Intrusion Detection System: A Survey. 2020, pp. 9-18.

23. Jung, E., Cho, I., & Kang, S. M. Моделювання агентів для подолання гетерогенності в IoT за допомогою шаблонів проектування. 2019, стор. 69-74.

24. Pajouh, H. H., Javidan, R., Khayami, R. and Ali, D. A Two-layer Dimension Reduction and Two-tier Classification Model for Anomaly-Based Intrusion Detection in IoT Backbone Networks. 2018, pp. 1-11.

25. Prabha, K. & Sree, S. S. Огляд методів і методів IPS. У International Journal of Computer Science Issues. 2017 р. С. 38-43.

26. Sherasiya, T. and Upadhyay, H. Intrusion Detection System for Internet of Things.2018, pp. 295-296.

27. Diro, A.; Chilamkurti, N.; Nguyen, V.D.; Heyne, W. A comprehensive study of anomaly detection schemes in iot networks using machine learning algorithms. Sensors 2021, pp.77-80.

28. Hasan, M.; Islam, M.M.; Zarif, M.I.I.; Hashem, M. Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. Internet Things 2019, pp.18-33.

ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ (Презентація)