НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

До захисту допущено
Завідувач кафедри

_____ **Дмитро ЛАНДЕ**
(підпис)

«_____»_____2025 р.

Дипломна робота

на здобуття ступеня бакалавра

за освітньо-професійною програмою «Системи, технології та математичні методи кібербезпеки»

спеціальності 125 «Кібербезпека»

на тему: Виявлення кібератак шляхом класифікації аномалій в показниках систем водопостачання

Виконав (-ла): здобувач вищої освіти **IV** курсу, групи ФБ-13
(шифр групи)

Буєва Христина Олександрівна
(прізвище, ім'я, по батькові) (підпис)

Керівник доцент кафедри ІБ, к.т.н., доцент, Стьопочкіна Ірина Валеріївна
(посада, науковий ступінь, вчене звання, прізвище, ім'я, по батькові) (підпис)

Рецензент доцент кафедри ММАД, к.т.н., доцент Лавренюк Алла Миколаївна
(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище, ім'я, по батькові) (підпис)

Засвідчую, що у цій дипломній роботі немає запозичень з праць інших авторів без відповідних посилань.
Здобувач вищої освіти _____
(підпис)

Київ – 2025 року

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

Рівень вищої освіти – перший (бакалаврський)
Спеціальність – 125 «Кібербезпека»
Освітньо-професійна програма «Системи, технології та математичні методи кібербезпеки»

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Дмитро ЛАНДЕ
(підпис)
«___»_____2025 р.

ЗАВДАННЯ
на дипломну роботу здобувачу вищої освіти

Буєвої Христини Олександрівни
(прізвище, ім'я, по батькові)

1. Тема роботи «Виявлення кібератак шляхом класифікації аномалій в показниках систем водопостачання»

керівник роботи Стьопочкіна Ірина Валеріївна к.т.н., доцент кафедри інформаційної безпеки,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
затверджені наказом по університету від «26» травня 2025 р. No 1761 - с.

2. Термін подання здобувачем вищої освіти роботи «13» червня 2025 р

3. Вихідні дані до роботи : Літературні джерела, дані про роботу стенда системи резервуарів в лабораторії ICS, датасет BATADAL.

4. Зміст роботи : Вибір датасета. Визначення характеристик, що можуть свідчити про вплив кіберфізичних атак. Ранжування ознак за важливістю. Виділення найбільш значущих ознак. Вдосконалення моделі класифікації з використанням штучних даних.

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо): презентація

6. Дата видачі завдання : 28 жовтня 2024 року

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів дипломної роботи	Примітка
1	Отримання завдання	11.10.2024 - 28.10.2024	виконано
2	Літературний огляд по темі	29.10.2024 - 20.01.2025	виконано
3	Аналіз наявних датасетів систем водопостачання	06.02.2025 - 21.02.2025	виконано
4	Вибір датасета для дослідження	21.02.2025 - 03.03.2025	виконано
5	Виділення характеристик для виявлення аномалій	21.04.2025 - 25.04.2025	виконано
6	Реалізація моделі класифікації	28.04.2025 - 01.05.2025	виконано
7	Практичний експеримент	02.05.2025 - 07.05.2025	виконано
8	Застосування техніки синтетичного збалансування датасета	08.05.2025 - 10.05.2025	виконано
9	Експериментальна оцінка ефективності запропонованої моделі класифікації та виділення найбільш значущих наборів ознак	12.05.2025 - 17.05.2025	виконано
10	Формування презентаційного матеріалу	20.05.2025 - 31.05.2025	виконано
11	Підготовка до передзахисту дипломної роботи	01.06.2025 - 12.06.2025	виконано
12	Підготовка до захисту дипломної роботи	13.06.2025 - 17.06.2025	виконано

Здобувач вищої освіти (підпис) Христина БУЄВА (Власне ім'я, ПРІЗВИЩЕ)

Керівник роботи (підпис) Ірина СТЬОПОЧКІНА (Власне ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ

Обсяг роботи 63 сторінки, 22 ілюстрації, 5 таблиць, 21 джерело літератури.

Об’єкт дослідження: аномалії в показниках систем водопостачання.

Предмет дослідження: методи виявлення аномалій, спричинених потенційним кіберфізичним впливом.

Мета : розвиток методів виявлення кіберфізичних атак на об'єкті критичної інфраструктури в області водопостачання.

Методи дослідження: аналіз, порівняння, моделювання, експеримент, статистичний аналіз.

Отримані результати: Сформовано перелік наявних датасетів систем водопостачання та проаналізовано їх зміст. Реалізовано декілька варіантів моделі класифікації з використанням методу машинного навчання Random Forest. Для обраного датасету сформовано шість груп ознак для виявлення кібератак. Застосовано техніку синтетичного збалансування SMOTE. За результатами практичного експерименту виділено найбільш значущі ознаки для виявлення аномалій та експериментально показано, що штучні дані дозволяють покращити якість класифікації.

Результати роботи були представлені на XXIII Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених.

Ключові слова: кібербезпека, об’єкти водопостачання, виявлення аномалій, кіберфізичні атаки

ABSTRACT

The volume of work is 63 pages, 22 illustrations, 5 tables, 21 sources of literature.

Object of research: anomalies in the indicators of water supply systems.

Subject of research: methods for detecting anomalies caused by potential cyber-physical impact.

Objective: development of methods for detecting cyber-physical attacks on critical infrastructure in the water supply sector.

Research methods: analysis, comparison, modeling, experiment, statistical analysis.

Results: A list of available datasets of water supply systems was formed and their content was analyzed. Several variants of the classification model using the Random Forest machine learning method were implemented. Six groups of features were formed for the selected dataset to detect cyberattacks. The SMOTE synthetic balancing technique was applied. According to the results of the practical experiment, the most significant features for detecting anomalies were identified and it was experimentally shown that artificial data allows improving the quality of classification.

The results of the work were presented at the XXIII All-Ukrainian Scientific and Practical Conference of Students, Postgraduate Students and Young Scientists.

Keywords: cybersecurity, water supply systems, anomaly detection, cyber-physical attacks

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ICS - Industrial Control System

IT - Information Technology

PLC - Programmable Logic Controller

SCADA - Supervisory Control and Data Acquisition

DCS - Distributed Control System

HMI - Human Machine Interface

SIS - Safety Instrumented System

RTU - Remote Terminal Unit

MTU - Master Terminal Unit

PERA - Purdue Enterprise Reference Architecture

IDS - Intrusion Detection System

IPS - Intrusion Prevention System

PCA - Principal Component Analysis

MAC - Media Access Control

IP - Internet Protocol

EDR - Endpoint Detection and Response

БД - База Даних

DoS - Denial of Service

DDoS - Distributed Denial of Service

FDI - False Data Injection

MITM - Man in the Middle

ВСТУП

Розвиток технологій та дедалі частіше впровадження автоматизованих систем керування сприяло збільшенню кількості кібератак, спрямованих на системи промислового управління (ICS). ICS є привабливою ціллю для злочинців, оскільки ці системи часто забезпечують функціонування критично важливої інфраструктури, як-от енергетика, транспорт, водопостачання та виробництво. Їх компрометація може щоразу спричиняти руйнівний вплив : значні економічні збитки, втрата робочих місць, зупинка роботи критично-важливих об’єктів і так далі. Розгляд існуючих методів детекції атак на ICS є безумовно актуальним, оскільки на сьогодні вже розроблено дуже багато систем виявлення кібератак, однак вони досі стикаються з складнощами і все ще не є довершеними [1].

Одним із підходів до виявлення кібератак є виявлення аномалій, які можуть сигналізувати про наявність кібернетичних втручань, які спрямовані на модифікацію показників системи, яка буде відрізнятись від нормальних значень. Таким чином зловмисник намагається дестабілізувати об’єкт. Часто такі впливи носять ситуативний характер (трапляються час від часу), або ж є доволі малими - що є серйозним викликом для відповідних систем виявлення аномалій. Таким чином, цікавим рішенням може бути пропозиція комбінованої системи - яка фіксуватиме відхилення від часового патерну стандартного перебігу процесу, або ж відхилення по абсолютних значеннях рівнів показників, або ж поступове плавне зменшення чи збільшення рівнів показників.

Метою дослідження є розвиток методів виявлення кіберфізичних атак на об'єктах критичної інфраструктури в області водопостачання.

Об’єктом дослідження є аномалії в показниках систем водопостачання.

Предметом дослідження є методи виявлення аномалій, спричинених потенційним кіберфізичним впливом.

Задачі дослідження:

1. здійснити аналіз існуючих рішень.
2. проаналізувати наявні моделі систем водопостачання, та наявність відповідних датасетів.
3. проаналізувати датасети з точки зору збалансованості, наявності необхідних характеристик (features). Виявити переваги та недоліки.
4. Обрати датасет, запропонувати перелік показників, обґрунтувати вибір методу штучного інтелекту для виявлення аномалій.
5. Розробити програмний застосунок для здійснення обчислювального експерименту.
6. Проаналізувати результати, з використанням існуючих лабораторних моделей - зробити висновки по перспективах виявлення аномалій запропонованим методом та інструментарем.

Методами дослідження є аналіз, порівняння, моделювання, експеримент, статистичний аналіз.

Інноваційність даної роботи полягає в тому, щоби провести відповідність між фізичними аномаліями та кібернетичними впливами, які їх потенційно могли спричинити. Робота заснована на сучасних підходах: методі Random Forest та синтетичному балансуванні датасету.

За результатами практичного експерименту виділено найбільш значущі ознаки для виявлення аномалій та експериментально показано, що штучні дані дозволяють покращити якість класифікації. Отримані результати дослідження в подальшому можуть використовуватись, як вихідні дані для системи прийняття рішень з метою визначення причини появи аномалії та коректного реагування на інцидент.

Результати роботи були представлені на XXIII Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених.

1 АНАЛІЗ ПРОБЛЕМИ ТА ОГЛЯД ПІДХОДІВ ДО ДЕТЕКЦІЇ АНОМАЛІЙ

1.1 ICS. Приклади кібератак на системи промислового управління

Промислові системи управління є програмованими системами, які використовуються для моніторингу, регулювання та управління важливими процесами промисловості, як от виробництво електроенергії, нафтогазоносні та транспортні системи тощо. Історично ICS не були наділені вбудованими функціями безпеки, оскільки вони були ізольованими від зовнішніх мереж. З розвитком та впровадженням технологій автоматизації та віддаленого керування, ICS починає використовувати системи інформаційних технологій. Саме це призводить до того, що раніше розповсюджені атаки в IT, почали з'являтися в промислових системах управління, через появу вразливостей типу незахищених протоколів, віддалених з'єднань тощо. Кібератаки на ICS чинять негативний, інколи нищівний, вплив на навколишнє середовище, суспільство та економіку [1].

У 2015 році, 23 грудня, відбулась кібератака на енергетичну мережу України, наслідком якої було виведення енергетичної системи з ладу. Шкідливе програмне забезпечення під назвою BlackEnergy, найбільше вразило "Прикарпаттяобленерго". Це спричинило перебої з електроенергією у близько чверті мільйона споживачів та відключення 30 підстанцій. Майже в той самий час зазнали атаки "Чернівціобленерго" та "Київобленерго". Тоді відбулось несанкціоноване втручання в роботу інформаційно-технологічної системи дистанційного доступу. Як наслідок, без електропостачання залишилося близько 80 тисяч споживачів, також було відключено 30 підстанцій, що забезпечували живлення кількох стратегічних об'єктів регіону. Дещо пізніше, у 2016 році, відбулася кібератака на компанію "Укренерго", а саме підстанцію "Північна".

Внаслідок атаки було знеструмлено північну частину правого берегу Києва, а також сусідні райони в області [2].

Ще одним відомим випадком є атака на ядерний об’єкт в Натанзі, Іран, 2011 року. Тоді було використано шкідливе програмне забезпечення, що має назву Stuxnet. Воно вважається першим шкідливим програмним забезпеченням, що було розроблене з ціллю атаки на обладнання ICS. Stuxnet заражає програмне забезпечення, призначене для моніторингу та запису даних до спеціальних програмованих логічних контролерів (Programmable Logic Controller). У разі успішної атаки на PLC, шкідливе програмне забезпечення могло контролювати з’єднання протоколу Profibus протягом 13 днів, коли уран додавався до центрифуг. Внаслідок атаки, в першу чергу, відбулося відключення функцій, що забезпечували коректне відключення системи, у разі несправності. Вважається, що це шкідливе програмне забезпечення спрямоване на прискорення темпів виходу з ладу обладнання, що в свою чергу призводить до більш високих експлуатаційних витрат через більш часту заміну обладнання [3].

У статті [4] детально описана відносно нещодавня кібератака, що відбулась у травні 2021 року, і вважається однією з найбільш руйнівних атак програм-вимагачів на Colonial Pipeline. Оскільки це одна з лідируючих нафтогазових компаній США, наслідки були серйозними як для економіки, так і для суспільства. Під час атаки злочинці отримали доступ до мережі Colonial через обліковий запис віртуальної приватної мережі, який використовував співробітник, що дозволило їм розгорнути програми-вимагачі та зашифрувати дані компанії. Внаслідок цієї атаки трубопровід вимушено тимчасово припинив свою роботу, що призвело до енергетичної кризи на Південному Сході Сполучених Штатів Америки, оскільки трубопровід забезпечував майже одну другу частину регіону.

1.2 Компоненти ICS

Поширеними типами ICS є диспетчерські системи контролю та збору даних (SCADA) та розподілені системи керування (DCS). Промислові системи управління працюють використовуючи відповідне апаратне та програмне забезпечення, протоколи зв'язку. Найпоширенішими типами компонентів ICS є програмовані логічні контролери (PLC), людино-машинний інтерфейс (HMI), різноманітні датчики, захисні системи (safety instrumented system, SIS), накопичувачі даних, блоки дистанційного керування (RTU), інженерні робочі станції. Саме середовище промислових систем управління часто може бути ієрархічним. Наприклад, у статті [1] наводиться пояснення усіх компонентів та їх графічне представлення у вигляді архітектури Purdue Enterprise Reference Architecture (PERA) - організація вище згаданих компонентів в шестирівневу архітектуру в межах конкретних мережевих зон.

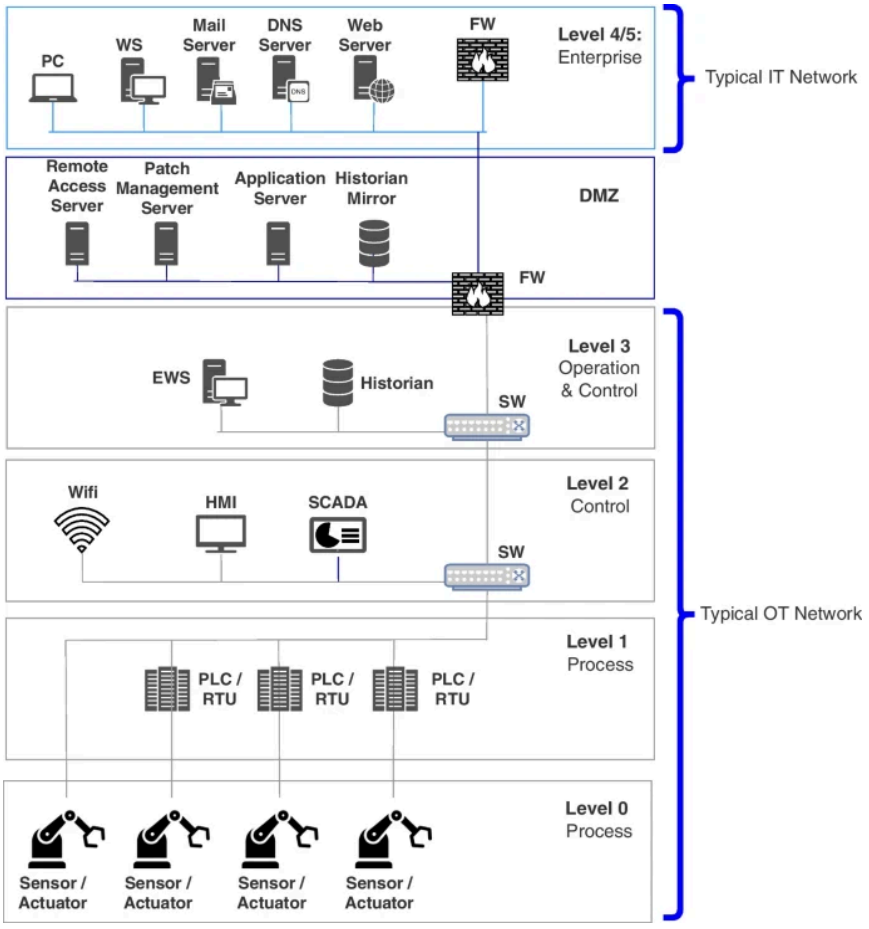

Рисунок 1.1 - Purdue Enterprise Reference Architecture [1]

Часто програмовані логічні контролери, що розташовані на першому рівні архітектури, стають ціллю кібератак. Зазвичай внаслідок таких атак відбувається пошкодження або модифікація даних, а також зміна конфігурацій. Яскравим прикладом є вже раніше описана кібератака з використанням шкідливого програмного забезпечення Stuxnet.

Людино-машинні інтерфейси зазвичай розташовані на другому рівні, забезпечують панель керування PLC, і працюють використовуючи операційні системи типу Windows. Вони є цікавими цілями для зловмисників, що знаходяться безпосередньо на об’єкті критичної інфраструктури.

Сенсори або виконавчі механізми розташовані на нульовому рівні. Їхня основна задача - збір та передача даних до PLC, а також керування механічними елементами системи. Відсутність на таких пристроях перевірки цілісності даних та процедури автентифікації робить їх вразливими до зміни логіки роботи.

Системи безпеки відповідають за автоматичне відключення процесів на об’єкті у разі виявлення загрози. Зазвичай вони контролюють граничні значення для процесів, щоб не було порушення визначених меж. Однак, системи безпеки можуть бути скомпроментовані і основна причина цьому - нестача моніторингу.

Ще одним компонентом ICS є системи, що зберігають дані про минулі події в базах даних для подальшого аналізу та створення звітів. За наявності накопичувачів даних, можна відстежити збої в роботі системи або ж виявляти аномалії.

Блоки дистанційного керування використовуються для моніторингу та контролю пристроїв, які розташовані у віддалених місцях, а також передачі даних до центральної станції. Ці блоки є вразливими до атак на протоколи зв’язку, перехоплення та модифікації даних.

Інженерні робочі станції використовуються інженерами для зміни або оновлення налаштувань різних систем ICS. Такі станції часто використовують застарілі операційні системи, що робить їх привабливою ціллю для кібератак.

1.3 Типові вразливості для мережевих протоколів ICS

Мережеві протоколи, що використовуються в промислових системах управління, зазвичай є протоколами передачі даних, що використовуються для комунікації між пристроями в системі. У статті [5] умовно ці протоколи класифікують за двома критеріями : чи є протокол специфічним для виробника і в якій галузі промисловості він використовується.

Таблиця 1.1 - Мережеві протоколи ICS [5]

Protocol	**Vendor-specific**	**Industrial sector**
Modbus	-	Process automation
Profinet/Profibus	-	Process automation
EtherCAT	-	Process automation
EtherNet/IP	-	Process automation
CIP	-	Process automation
Siemens S7	+	Process automation
Sinec H1	+	Process automation
DNP3	-	Process automation, power grid automation
ICCP	-	Power grid automation
BACnet	-	Building automation

Як видно з таблиці, більшість протоколів не є vendor-specific. Тобто при використанні цих протоколів в ICS можуть використовуватись пристрої різних компаній. Також більшість з наведених протоколів використовуються для

автоматизації процесів, деякі для автоматизації роботи електромереж та зчитування лічильників.

Розглянемо більш детально протоколи, які найчастіше використовуються в промислових системах управління [5].

Modbus (Modicon Communication Layer) - протокол прикладного рівня, розроблений Modicon (Schneider Electric). В його основі лежить використання архітектури “провідний-підлеглий” (master-slave). В ролі master може працювати HMI (human machine interface), а в ролі підлеглого будуть PLC (programmable logic controllers), або в ролі master - PLC, а в ролі slave - інші пристрої (датчики і так далі). Ідея полягає в тому, що головний надсилає запит підлеглому, очікує і приймає відповідь. Водночас протокол Modbus не використовує процедуру автентифікації та шифрування. Це може спричини до Denial-of-Service (DoS) атаки.

ICCP (Inter-Control Center Communication Protocol) також є протоколом прикладного рівня, що був розроблений для полегшення зв’язку між центрами управління електростанціями, забезпечивши обмін даними в реальному часі. В його основі лежить клієнт-серверна архітектура. Сервер містить дані та функції, а клієнт отримує доступ до них через запит. Під час процедури передачі даних, ICCP використовує двосторонню таблицю, яка виконує роль списку контролю доступу. В цьому протоколі також відсутні автентифікація та шифрування, що робить його вразливим до DoS атак.

BACnet (Building Automation and Control Networks) розроблений для об’єднання різних пристроїв, що вироблені різними компаніями, в одну систему. Протокол визначає набір з 25 стандартизованих типів об’єктів, що використовуються для взаємодії пристроїв всередині системи. Протокол є вразливим до атак типу DoS та MITM (Man-in-the-Middle).

DNP3 (Distributed Network Protocol) - розподілений мережевий протокол, що був розроблений в 1990 році для забезпечення передачі даних переважно в електроенергетичній промисловості. Цей протокол використовує архітектуру master-slave подібно до протоколу Modbus. Однак, на відміну від Modbus, у

протоколі DNP3 дозволяється двонапрямлена комунікація і “підлеглий” може надсилати повідомлення без запиту. Він також забезпечує механізм автентифікації, шифрування та часових міток, що допомагає запобігти атакам типу DoS, MITM, спуфінг.

1.4 Класифікація кібератак на ICS

У статті [6] пропонується поділ кібератак на промислові системи керування на чотири класи, а саме : розвідка (reconnaissance), ін’єкція відповіді та вимірювань (response and measurement injection), ін’єкція команд (command injection) та відмова в обслуговуванні (denial of service).

Розвідувальні атаки мають на меті збір інформації про систему керування, архітектуру її мережі. Також такі атаки можуть визначати різні характеристики пристроїв, наприклад : виробник, модель, номер партії, підтримувані мережеві протоколи, тощо. До методів атак розвідки можна віднести сканування адрес та кодів функцій, атаки на ідентифікацію пристроїв та сканування точок даних. Як результат, можна отримати інформацію, що дозволить створити певне уявлення про систему чи пристрій : який виробник пристрою, чим чи ким він використовується, з якою метою, тощо. Ця інформація може бути використана для виявлення вразливостей пристрою чи системи.

Зазвичай промислові системи керування працюють за такою схемою : клієнт надсилає запит, а сервер дає відповідь. Однак, як вже було акцентовано раніше, більшість протоколів передачі даних не мають механізму автентифікації, що дає змогу зловмисникам перехоплювати пакети, змінювати їх вміст, тощо. Атака ін’єкції відповіді може виникати, наприклад, в процесі контролю логічних програмованих контролерів або блоків віддаленого керування, може відбуватись захоплення мережевих пакетів та зміна їх вмісту, або ж створення стороннім девайсом фейкових відповідей клієнту.

Атаки типу командних ін’єкцій мають на меті впровадження помилкових команд чи зміну конфікурації. В промислових системах керування зазвичай такі атаки націлені на блоки дистанційного керування, які, як правило, запрограмовані на автоматичний моніторинг та керування процесами на певному віддаленому об’єкті. Потенційними наслідками впровадження помилкових команд зловмисником можуть бути : переривання процесу керування, ускладнення комунікації між пристроями, або повне її переривання, несанкціоновані зміни конфігурації та заданих значень процесів.

Ціллю атак класу DoS на ICS є спричинення неналежного функціонування певної її підсистеми, з метою фактичного виведення з ладу усієї системи керування. Такі атаки можуть бути реалізовані методом виведення з ладу програм керування, або ж можуть бути проведені навіть фізичним шляхом : це, наприклад, включає ручне навмисне відкриття чи закриття клапанів, увімкнення чи вимкнення перемикачів, руйнування обладнання, тощо.

1.5 Огляд існуючих підходів до детекції аномалій з використанням методів машинного навчання

У статті [7] пропонується до розгляду певний перелік методів детекції аномалій в системах промислового контролю з використанням методів машинного навчання.

Основний фокус дослідження спрямований на системи виявлення вторгнень (IDS) у SCADA-системах, які контролюють фізичні процеси за допомогою даних з сенсорів та виконують моніторинг показників роботи об’єктів критичної інфраструктури.

Методи машинного навчання, що були розглянуті, можна поділити на дві групи : контрольовані та неконтрольовані. Контрольовані методи навчаються на основі наявних міток - розмітка даних на нормальну поведінку та аномалію. Такі

методи машинного навчання використовуються для завдань класифікації та регресії. Неконтрольовані методи самостійно аналізують структуру, не знаючи, які дані є нормальними, а які аномальними, і використовуються для кластеризації.

До контрольованих методів, тобто supervised learning, можна віднести логістичну регресію, KNN, SVM, Decision Tree, J48, Random Forest, Naive Bayes. До unsupervised learning, неконтрольованих методів, було віднесено кластеризацію k-means (групує схожі дані для виявлення аномалій) та k-medoids (цей підхід є менш чутливим до шуму), а також аналіз головних компонент - зменшує кількість характеристик до розгляду (PCA) .

У статті також пропонується використання гібридних підходів до виявлення аномалій в показниках роботи систем об’єктів критичної інфраструктури. Використання методу J48 у поєднанні з Bayes Network забезпечує кращу продуктивність класифікації аномалій. Поєднання логістичної регресії з Bayes Network використовує як ймовірнісні, так і регресійні підходи.

Перелік методів, що базуються на поведінковому аналізі, включає інспекцію пакетів (аналіз потоку даних між компонентами системи промислового контролю) та профілювання мережевого трафіку (визначає нормальну поведінку та виявляє аномалії)

Окремо зазначається можливість використання глибоких нейронних мереж (ANN) та ансамблевого методу машинного навчання Random Forest для підвищення точності роботи моделі.

Висновки до Розділу 1

У першому розділі було розглянуто одні з найвідоміших випадків кібератак на об’єкти критичної інфраструктури в Україні та світі. Було розглянуто поняття системи промислового контролю (ICS) та їхні типи.

Описано такі компоненти ICS, як програмовані логічні контролери, людино-машинний інтерфейс, різноманітні датчики, захисні системи, накопичувачі даних, блоки дистанційного керування, інженерні робочі станції. Також проаналізовано організацію компонент у шестирівневу архітектуру PERA.

Розглянуто найбільш поширені мережеві протоколи, що використовуються в промислових системах управління для комунікації між пристроями в системі.

Було описано загальну класифікацію кібератак на ICS, а також проведено огляд існуючих підходів до детекції аномалій в показниках роботи об’єктів критичної інфраструктури з використанням методів машинного навчання.

2 ВИБІР МЕТОДУ РОЗВ'ЯЗАННЯ ЗАДАЧІ

2.1 Вибір сфери застосування

ICS використовуються в різних сферах критичної інфраструктури. Серед них можна виділити водопостачання та водовідведення, газопостачання, електроенергетика, нафтопереробна та хімічна промисловості, транспорт, харчова промисловість та виробництво. Як показав проведений аналіз літератури, наразі не існує ідеальної системи промислового контролю, яка б не мала вразливостей і, більш того, була універсальною.

Системи промислового контролю є дуже різноманітними. Можна навіть сказати, що кожна з них є унікальною, оскільки адаптована не лише під особливості конкретного сектору інфраструктурного забезпечення, а й забезпечує контроль певним об'єктом чи технічним комплексом, включаючи особливості будови, мережі, діяльність та багато інших специфічних факторів.

Для розуміння певної ICS-системи, потрібно не лише проаналізувати її компоненти та взаємозв'язки між ними, а, перш за все, вивчити сферу діяльності та дослідити увесь комплекс технологічних та експлуатаційних процесів об'єкта певної промислової системи.

Для проведення якісного дослідження варто визначитись з конкретною сферою, яка буде розглядатись. Проаналізувавши літературні джерела, була обрана сфера водопостачання.

Системи забезпечення водопостачання та водовідведення є критично важливими для забезпечення життєдіяльності суспільства та збереження екологічного середовища. Будь-які збої в роботі систем водопостачання та водовідведення можуть мати катастрофічні наслідки : від масових отруєнь до соціальних криз. Системи водопостачання є вразливими не лише до кібератак, а й до фізичного впливу, наприклад, забруднення. Аномалії серед показників якості

води можуть тривалий час залишатись непоміченими без спеціального моніторингу. Відхилення в роботі приладів, таких як помпи, насоси, резервуари, регулятори та клапани, прилади для вимірювання та контролю і так далі, можуть спричини аварії на критичних об’єктах водозабезпечення. Дослідження різного роду аномалій може допомогти виявити інциденти та дисфункції в роботі об’єкта та вчасно запобігти негативним наслідкам, які часто можуть стати критичними для навколишнього середовища та суспільства.

2.2 Наявні моделі водопостачання. ASTANK 2

2.2.1 Загальний опис архітектури ASTANK2

У статті [8] детально описано систему резервуарів ASTANK2. ASTANK2 - це лабораторна гідравлічна модель, яка складається з двох резервуарів різної геометричної форми. Система має широкий спектр застосування в освітніх та дослідницьких цілях : використовується в галузі автоматизації, моделюванні та контролі систем, а також сприяє кращому розумінню гідравлічних процесів. ASTANK2 має можливість моделювання різних сценаріїв роботи гідравлічної системи, включаючи вплив збурень та зміну конфігурацій, що дозволяє тестувати алгоритми для забезпечення контролю систем.

Лабораторна гідравлічна установка є багатокомпонентною. Усі її складові забезпечують функціональність моделі та дають змогу моделювати гідравлічні процеси.

ASTANK2 складається з резервуарів, насосної системи, датчиків та контролерів, електроклапанів - кожен із цих компонентів можна змінювати та налаштовувати окремо. Розглянемо детальніше архітектуру системи ASTANK2, яка зображена на рисунку :

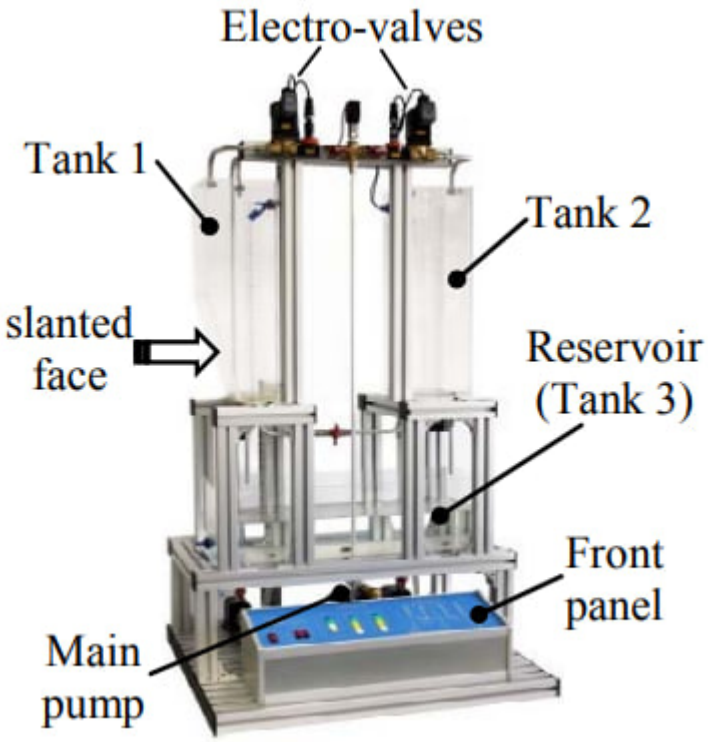

Рисунок 2.1 - Архітектури системи ASTANK2 [9]

Система складається з трьох резервуарів : двоє розміщені ліворуч (Tank 1) та праворуч (Tank 2), один знизу (Tank 3). Резервуар 1 має одну нахилену стінку. Завдяки цій особливості зміна рівня води в першому резервуарі нелінійно впливає на зміну об’єму. Це є ключовою характеристикою цієї модельованої системи, оскільки нахилена стінка ускладнює математичну модель системи та дозволяє більш реалістично змоделювати процеси. Резервуар 2 має звичайну форму паралелепіпеда, без додаткових факторів забезпечення нелінійності. Резервуар 3 є накопичувальним. Він забезпечує безперервну циркуляцію води в системі - забезпечення водою першого та другого баків через насосну систему. Регулювання потоку води, що подається до Tank 1 та Tank 2 відбувається за допомогою електроклапанів.

2.2.2 Розгляд всіх компонентів системи

Нижче представлено схему моделі ASTANK2, включаючи всі її елементи :

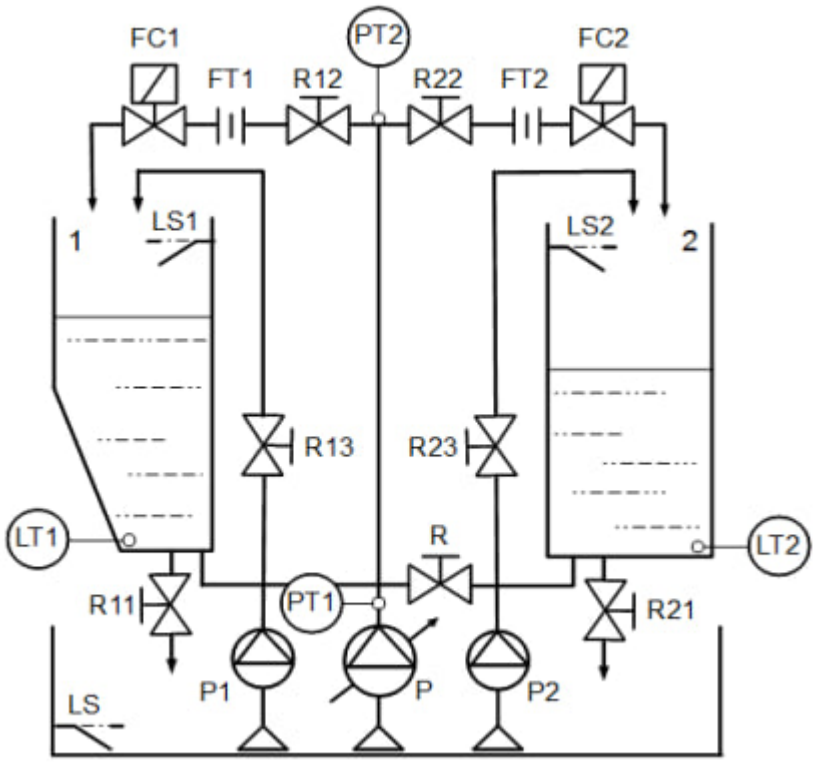

Рисунок 2.2 - Схема моделі ASTANK2 [8]

Вода з накопичувального резервуару (Tank 3), її рівень вимірюється датчиком LS, подається до вертикальної труби за допомогою головного насоса P. Тиск в цій трубі контролюється датчиками PT1 та PT2. Далі потік води розподіляється між резервуарами через горизонтальну трубу. R12 та R22 - крани для контролю наповнення резервуарів 1 та 2 відповідно. Електроклапани FC1 та FC2 регулюють потоки води спрямовані до резервуарів, а датчики витрат FT1 та FT2 вимірюють вхідні потоки. В резервуарах 1 та 2 встановленні датчики переповнення LS1 і LS2 відповідно. Контроль рівня води в Tank 1, Tank 2 відбувається за допомогою датчиків LT1, LT2. За допомогою крану R можна регулювати режим роботи моделі :

1. Незалежні резервуари. Кран R закритий, потік подається до кожного резервуара окремо.
2. Резервуари пов’язані один з одним. У такому випадку, кран R відкритий і резервуари є об’єднані через нижній трубопровід, тобто рівень води вирівнюється.

В системі також наявні додаткові насоси P1 та P2, які можуть подавати воду до резервуарів 1 та 2, через вертикальні трубопроводи. Таким чином можна змоделювати випадкові збурення в системі, створюючи коливання рівнів води, які можна додатково контролювати ручними кранами R13 і R23. Крани R11 і R22 забезпечують злив води з резервуарів 1 та 2 відповідно.

Можна виділити декілька ключових параметрів, від яких залежить рівень води в усіх трьох резервуарах :

1. Напруга на електроклапанах FC1, FC2. Напруга, що подається на електроклапан визначає міру його відкриття, тобто пропускну здатність. Чим більша напруга - тим більший потік спрямовано до резервуару.
2. Напруга на головному насосі P - впливає на загальний об’єм води, поданої до резервуарів 1 та 2.
3. Режим роботи додаткових насосів P1, P2. Стан насосів (ON/OFF) регулює наявність збурень.

2.2.3 Поділ системи на блоки та детальний опис їхньої роботи

Для покращення розуміння, модель ASTANK2 також можна розглянути як систему кількох лінійних і нелінійних блоків (живлення, електроклапанів, резервуарів), що зображено на рисунку :

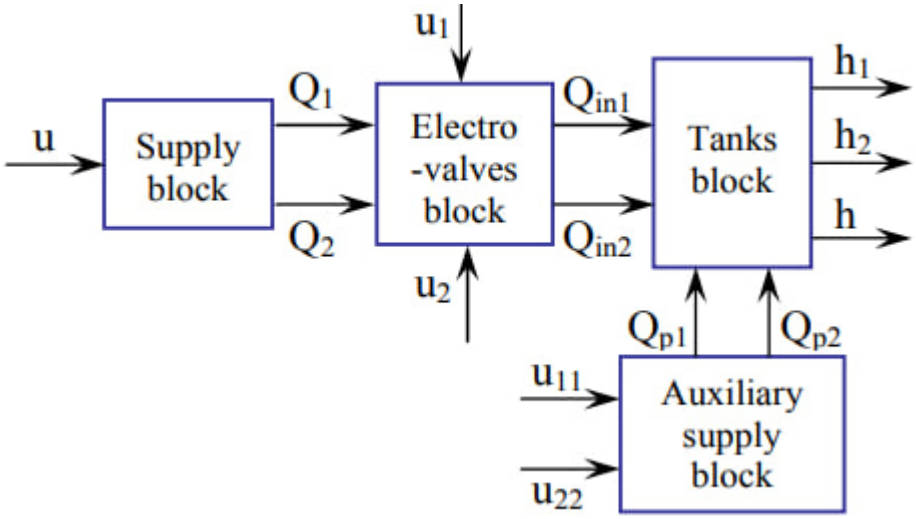

Рисунок 2.3 - Модель ASTANK2 у вигляді лінійних і нелінійних блоків [8]

Більш детально загальна схема роботи моделі ASTANK2 представлена на рисунку :

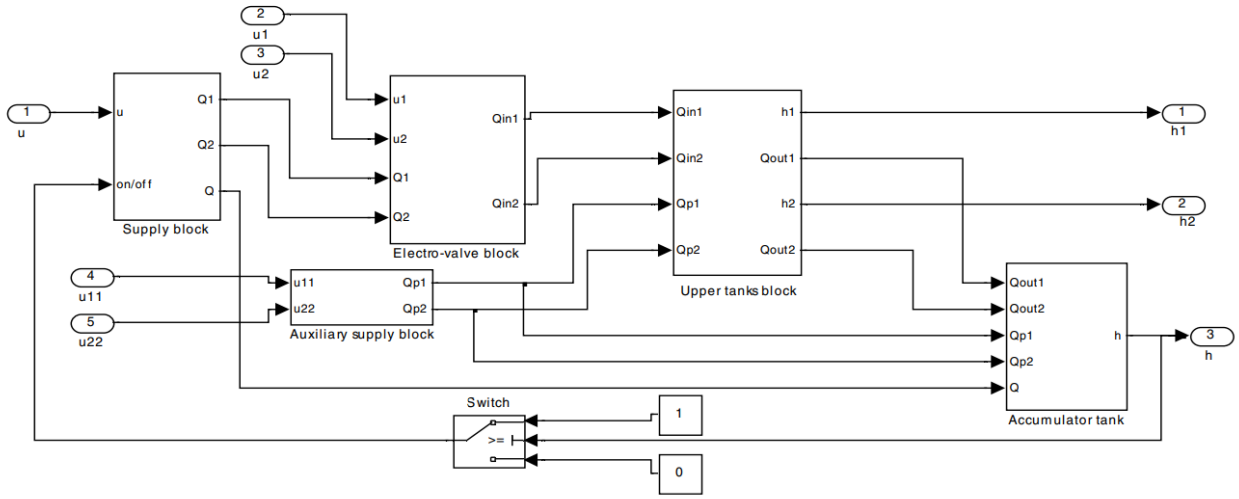

Рисунок 2.4 - Схема роботи моделі ASTANK2 [8]

Блок живлення включає в себе насосну систему та контролює її роботу. Зі схеми видно, що на вхід приймається сигнал u - це значення напруги в насосі. На виході отримуємо два сигнали Q_1 та Q_2 - це потоки води, спрямовані до електроклапанів FC1, FC2 відповідно. Допоміжний блок живлення відповідає за допоміжну насосну систему. Він отримує сигнали u_{11} та u_{22} - керуючі сигнали для додаткових насосів P1 та P2 зі значенням стану ON/OFF. Далі, якщо керуючий сигнал має значення ON, до резервуару подається додатковий потік води (збурення). На схемі ці потоки позначені як Q_{p1} та Q_{p2} для Tank 1 і Tank 2 відповідно.

Аналітична модель блоку живлення, базується на передаточних функціях датчиків тиску та витрати і на математичній моделі процесу течії потоку води через верхню горизонтальну трубу, між точкою розгалуження та електроклапанами. Основними аспектами цього блоку є підсилення датчика тиску $(K_p = 1)$ та підсилення датчиків витрати $(K_Q = 7.63)$ із часовою затримкою $T_Q = 1$ секунда. Передаточна функція датчиків витрати має вигляд : $H(s) = 7.63 * e^{-s}$.

Процес течії потоку води в цьому блоці описується рівнянням :

$$\rho L_p \cdot \frac{dQ_{1,2}(t)}{dt} + \rho \cdot S \cdot Q_{1,2}(t)^2 = \Delta P \cdot \alpha \cdot S$$, де

ΔP - вхідний сигнал диференціального тиску;

$Q_{1,2}(t)$ - потік води в лівій та правій частині верхньої горизонтальної труби;

ρ - густина води;

S - площа поперечного перерізу труби (0.00007854 м^2);

L_p - довжина труби (0.04 м);

α - коефіцієнт витрати (0.127).

Вхідні сигнали для блоку електроклапанів : Q_1 та Q_2; u_1 та u_2 - значення напруги, котра подається до FC1 і FC2 відповідно. Сигнали u_1 та u_2 визначають пропускну здатність електроклапанів. Вихідними сигналами для цього блоку є Q_{in1} та Q_{in2}, що означають реальні потоки води, які надходять до резервуарів 1 та 2.

Блок електроклапанів моделюється нелінійним рівнянням, що враховує співвідношення між вхідним та вихідним потоком та пропускною здатністю електроклапана :

$$T_{EV} \cdot \frac{dQ_{out}(t)}{dt} + Q_{out}(t) = \gamma(u_{EV}) \cdot Q_{in}(t)$$, де

$Q_{out}(t)$ - вихідний потік;

$Q_{in}(t)$ - вхідний потік;

$\gamma(u_{EV})$ - коефіцієнт потоку, що нелінійно залежить від значення напруги;

T_{EV} - затримка часу роботи електроклапана.

Враховуючи описані раніше вхідні та вихідні сигнали для блоку електроклапанів, загальне рівняння можна адаптувати для кожного електроклапана (FC1, FC2) :

$$T_{EV} \cdot \frac{dQ_{\{in1,\, in2\}}(t)}{dt} + Q_{\{in1,\, in2\}}(t) = u_{\{1,2\}}(t) \cdot Q_{\{1,2\}}(t)$$, де

$T_{EV} = 0.0125$ секунд (визначений час затримки для моделі електроклапанів Burkert).

Блок резервуарів включає два верхніх резервуари та один накопичувальний, що знаходиться внизу. Вхідними сигналами для цього блоку є :

- Q_{in1} та Q_{in2} - потоки води, що надходять до першого та другого резервуару відповідно, через головну насосну систему;
- Q_{p1} та Q_{p2} - потоки води, що надходять до першого та другого резервуару відповідно, через додаткову насосну систему.

На виході, ми отримуємо три сигнали, а саме : h - рівень води в накопичувальному резервуарі, h1 та h2 - рівень води в резервуарі 1 та 2.

Лівий резервуар (Tank 1) має цікаву геометричну форму, а саме одну нахилену стінку, як зображено на рисунку. Основними параметрами резервуара є : L - довжина дна, l - ширина дна, H - максимальна висота нахиленої стінки, θ - кут нахилу стінки.

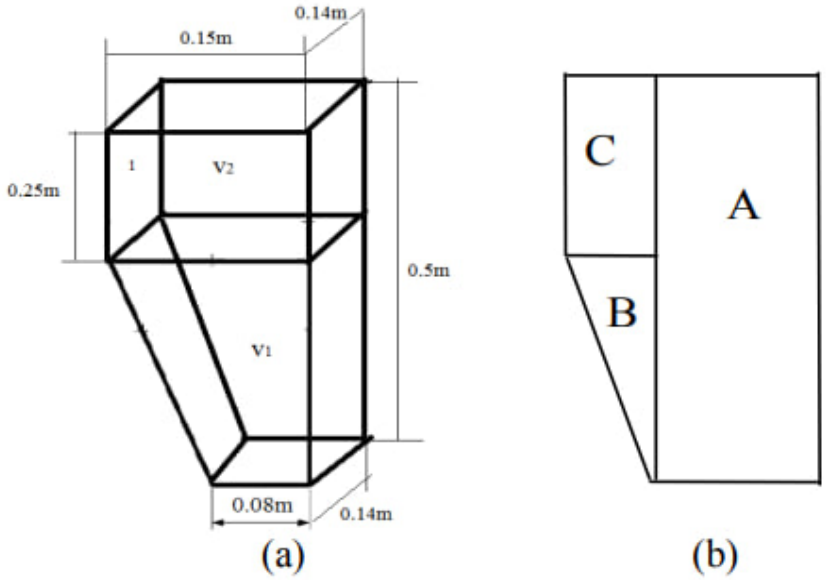

Рисунок 2.5 - Особливості будови лівого резервуара [8]

Якщо резервуари працюють незалежно один від одного, то його математична модель базуватиметься на рівнянні, як залежність об’єму від різниці між вхідним і вихідним потоком :

$$\frac{dV_1(t)}{dt} = Q_{in1}(t) - Q_{out1}(t)$$

Однак, вихідний потік Q_{out} розраховується з урахуванням гідростатичного тиску, з яким вода діє на дно резервуара. Значення тиску залежить від поточного об'єму води та геометрії резервуара. Саме тому маємо поділ резервуара 1 на три зони, які зображені на рисунку (b).

Об'єм резервуара 1 визначається, як сума об'ємів кожної зони. Для зони А об'єм розраховується за формулою :

$$V_A = L \cdot l \cdot h_1$$

Для зони В ($h_1 \leq H$) :

$$V_B = \frac{1}{2} \cdot l \cdot h_1^2 \cdot tan\theta$$

Об'єм зони С ($h_1 > H$) :

$$V_C = l \cdot (h_1 - H) \cdot H \cdot tan\theta$$

Значення гідростатичного тиску теж залежить від конкретної зони. У випадку, коли $h_1 \leq H$, гідростатичний тиск змінюється квадратично, залежно від рівня води в резервуарі :

$$P(h_1) = \rho \cdot g \cdot h_1 \cdot (1 + \frac{h_1}{4L} sin(2\theta))$$

Якщо $h_1 > H$, то маємо лінійну залежність :

$$P(h_1) = \rho \cdot g \cdot h_1 \cdot (1 + \frac{H}{2L} sin(2\theta)) - \rho \cdot g \cdot \frac{H^2}{2L} sin(2\theta)$$

Вихідний потік Q_{out1} розраховується за формулою : $Q_{out1} = S_1 \cdot v_1 \cdot \alpha_1$, де S_1 - площа поперечного перерізу дренажної труби,

v_1 - швидкість потоку, α_1- коефіцієнт витоку (зазвичай встановлюється значення $\alpha_1 = \alpha_1/S_1$). Рівняння Бернуллі дозволяє визначити швидкість потоку через перепад тиску : $v_1 = \sqrt{\frac{2\Delta P_1}{\rho}}$, де ΔP_1- диференціальний тиск між водою в резервуарі та зовнішнім середовищем. Таким чином, маємо оновлену формулу для вихідного потоку Q_{out1} :

$$Q_{out1} = \alpha_1 \cdot \sqrt{\frac{2\Delta P_1}{\rho}}$$

Тепер можна записати більш детальну математичну модель, що базується на рівнянні, як залежність об’єму від різниці між вхідним і вихідним потоком :

якщо $h_1 \leq H$:

$$(L \cdot l + l \cdot tan\theta \cdot h_1(t))\frac{dh_1(t)}{dt} =$$

$$= Q_{in1}(t) - \alpha_1\sqrt{2 \cdot (g \cdot h_1(t) \cdot (1 + \frac{h_1(t)}{4L}sin(2\theta)))}$$

якщо $h_1 > H$:

$$l(L + H \cdot tan\theta)\frac{dh_1(t)}{dt} =$$

$$= Q_{in1}(t) - \alpha_1\sqrt{2 \cdot g \cdot h_1(t) \cdot (1 + \frac{H}{2L}sin(2\theta)) - g \cdot \frac{H^2}{2L}sin(2\theta)}$$

Для резервуара 2 все має дещо простіший вигляд, оскільки він має форму звичайного паралелепіпеда.. У правому резервуарі залежність об’єму від висоти можна записати у вигляді рівняння:

$\frac{dV_2(t)}{dt} = Q_{in2}(t) - Q_{out2}(t)$, де $Q_{out2}(t) = a_2\sqrt{2gh_2}$

Таким чином рівняння процесу наповнення або дренажу для резервуара 2 має вигляд :

$$A_2\frac{dh_2(t)}{dt} = Q_{in2}(t) - a_2\sqrt{2gh_2(t)}, \text{ де}$$

A_2- площа поперечного перерізу правого резервуару.

Третій резервуар забезпечує водою систему, а також наповнюється завдяки дренажним потокам з резервуарів 1 та 2. Зміну об’єма води в баці можна описати рівнянням, як залежність об’єму від потоків, що надходять з лівого та правого резервуарів, а також потоку, що подається до системи через головний трубопровід ($Q(t)$) :

$$\frac{dV(t)}{dt} = Q_{out1}(t) + Q_{out2}(t) - Q(t)$$

Резервуари 1 та 2 можуть працювати поєднано у випадку, коли відкрито кран R, як вже зазначалося раніше. У такому разі, потік між лівим та правим резервуаром описується за допомогою рівняння :

$$Q_c = a_c \cdot sign(P(h_1) - \rho g h_2)\sqrt{\frac{2(P(h_1)-\rho g h_2)}{\rho}}$$, де

a_c - константа дренажу (залежить від крану R);

$sign(P_1 - P_2)$ - функція, що визначає напрямок потоку, який виникає завдяки різниці тисків між баками. Ця функція повертає значення 1, якщо $P_1 > P_2$ (тобто потік спрямований з резервуара 1 до резервуара 2);

-1, якщо $P_1 < P_2$; 0, якщо $P_1 = P_2$

Детально розібравши особливості роботи кожного з трьох резервуарів, аналітична модель роботи всієї системи матиме вигляд :

$$\frac{dV_1(t)}{dt} = Q_{in1}(t) + Q_{p1}(t) - Q_{out1}(t) - Q_c(t)$$

$$\frac{dV_2(t)}{dt} = Q_{in2}(t) + Q_{p2}(t) - Q_{out2}(t) + Q_c(t)$$

$$\frac{dV(t)}{dt} = Q_{out1}(t) + Q_{out2}(t) - Q(t) - Q_{p1}(t) - Q_{p2}(t)$$

2.3 Вибір датасету

Датасет є основою для проведення якісного та результативного дослідження з виявлення аномалій. Для того, щоб отримати наближені до реальності результати, обраний набір даних повинен в достатній кількості містити збалансовані та різноманітні дані, з яких можна виокремити корисну для дослідження інформацію.

Є багато даних для сфери водопостачання. Одними з найбільш відомих та найбільш популярних датасетів є SWaT, HAI, IUNO, WST, WADI, Festo, BATADAL.

Secure Water Treatment (SWaT) датасет містить дані, за період одинадцяти днів, з яких протягом семи днів збиралися нормальні дані, а протягом чотирьох - застосовувались сценарії атак. Датасет містить дані з 51 датчиків та виконавчих механізмів, а також відповідні відмітки для цілком нормальної та аномальної поведінки системи. Впродовж чотирьох днів було здійснено 41 атаку, відповідно до застосованих моделей атак [10].

HIL-based Augmented ICS Security (HAI) - набір даних отриманий з моделі системи промислового управління, доповненої симулятором Hardware-in-the-loop (HAI), що описує виробництво електроенергії з використанням парових турбін і гідроакумулюючих гідроелектростанцій [11].

Наступний датасет, що має назву Water Storage Tank and Gas Pipeline SCADA systems (WST) був сформований з двох лабораторних SCADA систем : водного резервуару та газопроводу. Ці набори даних включають мережевий трафік, результати процесу контролю та різні вимірюванні дані датчиків [12].

Water Distribution Testbed (WADI) містить дані зі зменшеної лабораторної версії системи розподілу міського водопостачання. Також підключений до випробувальних стендів очищення води, виробництва та розподілу електроенергії [13].

The BATtle of the Attack Detection ALgorithms [14] (BATADAL) містить дані водорозподільної системи C-Town. C-Town містить 388 вузлів, з'єднаних 429 трубами, і поділений на 5 зон вимірювання району (DMA). Дані включають рівень води у всіх 7 резервуарах мережі (T1–T7), стан і потік усіх 11 насосів (PU1–PU11) та одного керованого клапана (V2), а також тиск у 24 трубах, що відповідає значенням на вході та виході насосів і клапана. Водорозподільчу систему C-Town представлено на рисунку 2.6 :

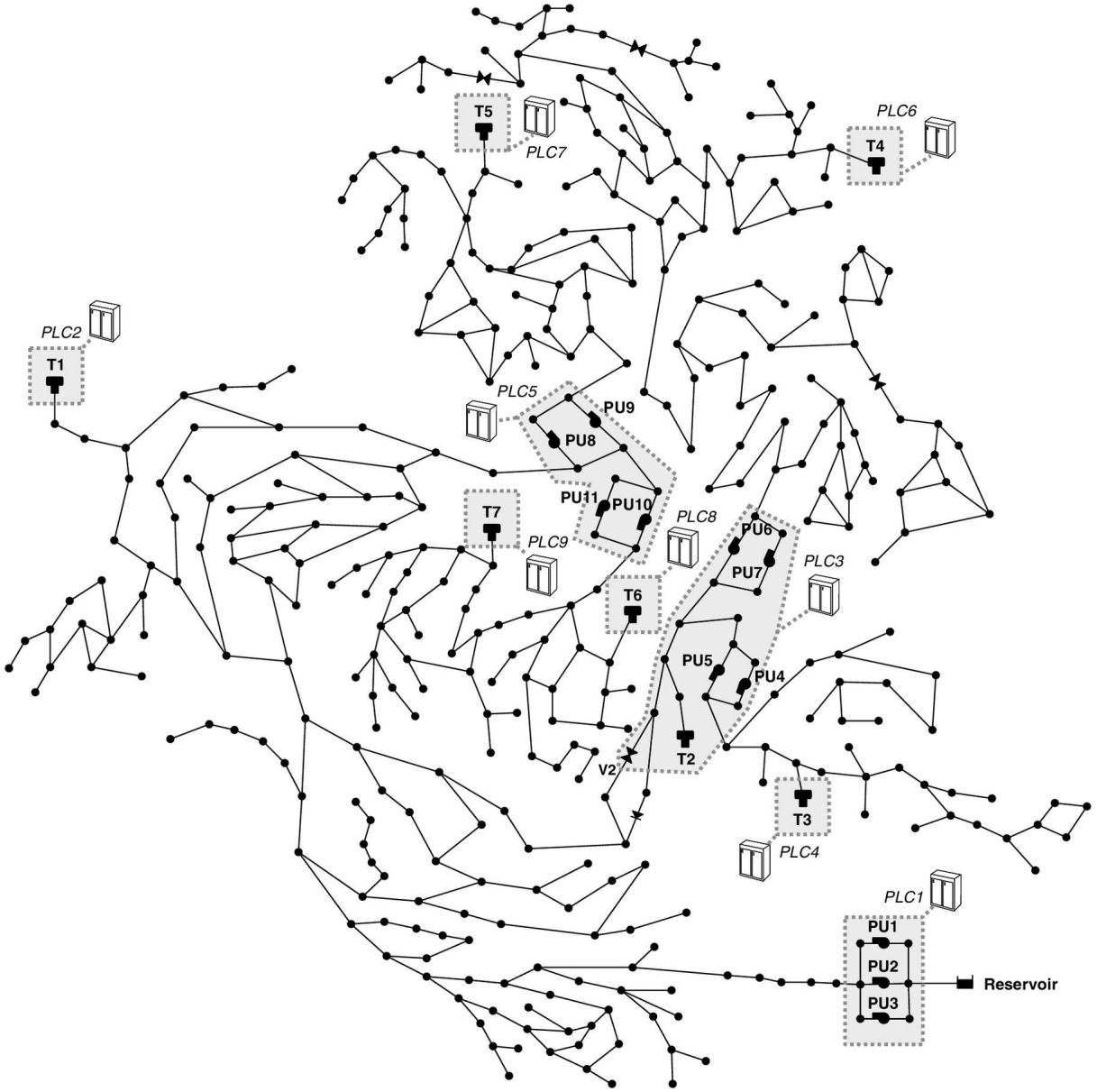

Рисунок 2.6 - Водорозподільча система C-Town [14]

Набір даних складається з двох навчальних датасетів та одного тестового. Датасети містять 43 стовпці, а також стовпець з мітками 1 або 0 (нуль відповідає нормальній поведінці системи, а одиниця свідчить про те, що система піддається впливу якоїсь атаки).

У першому навчальному датасеті знаходяться дані нормальної роботи системи, що описують роботу системи протягом року. У другому навчальному датасеті позначено приблизно 7 атак і ці дані збирались протягом 6 місяців. Детальніше описано на рисунку 2.7 :

Attacks featured in Training dataset 2

ID	Starting time [dd/mm/YY HH]	Ending time [dd/mm/YY HH]	Duration [hours]	Attack description	SCADA concealment	Labeled [hours]
1	13/09/2016 23	16/09/2016 00	50	Attacker changes L_T7 thresholds (which controls PU10/PU11) by altering SCADA transmission to PLC9. Low levels in T7.	Replay attack on L_T7 .	42
2	26/09/2016 11	27/09/2016 10	24	Like Attack #1.	Like Attack #1 but replay attack extended to PU10/PU11 flow and status.	0
3	09/10/2016 09	11/10/2016 20	60	Attack alters L_T1 readings sent by PLC2 to PLC1, which reads a constant low level and keeps pumps PU1/PU2 ON. Overflow in T1.	Polyline to offset L_T1 increase.	60
4	29/10/2016 19	02/11/2016 16	94	Like Attack #3.	Replay attack on L_T1, PU1/PU2 flow and status, as well as pressure at pumps outlet.	37
5	26/11/2016 17	29/11/2016 04	60	Working speed of PU7 reduced to 0.9 of nominal speed causes lower water levels in T4.		7
6	06/12/2016 07	10/12/2016 04	94	Like Attack #5, but speed reduced to 0.7.	L_T4 drop concealed with replay attack.	73
7	14/12/2016 15	19/12/2016 04	110	Like Attack #6.	Replay attack on L_T1, as well as PU1/PU2 flow and status.	0

Рисунок 2.7 - Опис кібератак для другого навчального датасета BATADAL [14]

Тестовий набір даних збирався протягом трьох місяців. На відміну від навчальних даних, у тестовому датасеті аномальні дані, що свідчать про атаку, не є позначеними. Детальний список атак у тестовому наборі даних представлено на рисунку 2.8 :

Attacks featured in the Test dataset

ID	Starting time [dd/mm/YY HH]	Ending time [dd/mm/YY HH]	Duration [hours]	Attack description	SCADA concealment
8	16/01/2017 09	19/01/2017 06	70	Attacker changes L_T3 thresholds (which control PU4/PU5) by gaining control of PLC3. Low levels in T3.	Replay attack on L_T3, as well as PU4/PU5 flow and status.
9	30/01/2017 08	02/02/2017 00	65	Attack alters L_T2 readings arriving to PLC3, which reads a low level and keeps valve V2 OPEN, leading T2 to overflow.	Polyline to offset L_T2 increase.
10	09/02/2017 03	10/02/2017 09	31	Malicious activation of pump PU3	
11	12/02/2017 01	13/02/2017 07	31	Similar to Attack #10	
12	24/02/2017 05	28/02/2017 08	100	Similar to Attack #9	Replay attack on L_T2, V2 flow and status, as well as V2 inlet + outlet pressure readings (P_J14, P_J422)
13	10/03/2017 14	13/03/2017 21	80	Attacker changes L_T7 thresholds (which control PU10/PU11) by gaining control of PLC5, causing the pumps to switch ON/OFF continuously.	Replay attack on L_T7, PU10/PU11 flow and status, as well as inlet + outlet pressure readings (P_J14, P_J422). Inlet pressure concealment terminates before that of other variables.
14	25/03/2017 20	27/03/2017 01	30	Alteration of T4 signal arriving to PLC6. Overflow in T6.	

Рисунок 2.8 - Опис кібератак для тестового датасета BATADAL [14]

Набір датасетів BATADAL було обрано для подальшого проведення дослідження, оскільки він є загальнодоступним, містить різноманітні дані, був створений для перевірки роботи алгоритмів виявлення атак на водорозподільні системи.

2.4 Ознаки для детекції аномалій

Для ефективного виявлення аномалій в роботі різних систем, перш за все потрібно розуміти, де їх шукати та на які дані звертати увагу. Для сфери водопостачання та водозабезпечення можна виокремити декілька загальних характеристик, так званих features, для виявлення аномалій :

- зміна гідравлічних параметрів : до цієї категорії належать такі показники, як зміна рівня води в резервуарах або зміна характеристик потоку води в трубах (наприклад, тиск, швидкість потоку, об’єм та рівень води);
- стан компонентів системи : сюди відноситься несанкціоноване відкриття або закриття шлюзів та клапанів, а також вмикання або вимикання датчиків;
- зміна граничних значень;
- розбіжність між даними з різних сенсорів (наприклад, ми бачимо, що через певний клапан до резервуару було подано великий об’єм води, а датчик рівня фіксує мінімальний об’єм);
- часові ряди процесів;
- зміна фізичних, біологічних та хімічних характеристик води.

Наприклад, для гідравлічної моделі ASTANK2 можна виокремити набір показників, дані яких можуть бути використані для детекції аномалій:

Таблиця 2.1 - Ознаки для детекції аномалії в показниках системи резервуарів ASTANK2 [8]

Показник	Опис
LT1, LT2	вимірювачі рівня води в резервуарах 1 та 2
LS1, LS2	датчики переповнення в резервуарах 1 та 2
LS	датчик для контролю нижньої межі рівня води в резервуарі 3
P	основний насос
P1, P2	додатковий насос
PT1, PT2	датчики тиску (також PT2 може надсилати сигнал про недостатню кількість води чи тиску для одночасного забезпечення двох гілок трубопроводу)
FT1, FT2	датчики витрат, які вимірюють вхідні потоки до резервуарів 1 та 2
FC1. FC2	електроклапани, що регулюють вхідні потоки до резервуарів 1 та 2
R13, R23	крани для регулювання збурень
R11, R12	крани для зливу води з резервуарів 1 та 2

Якщо розглянути представлені вище компоненти та врахувати особливості роботи системи, то можна зробити висновок, що усі вони залежать від :

- напруги, яка подається до електроклапанів;
- напруги, яка подається до головного насосу;
- стану додаткових насосів.

Для датасетів BATADAL можна виокремити наступний набір ознак для виявлення аномалій [14] :

- L_T* - рівень води в резервуарах;
- S_PU*- стан насоса;
- S_V* - стан клапана;
- F_PU*- витрата насоса
- F_V* - витрата клапана;
- P_J* - значення тиску.

2.5 Вибір методу розв'язання задачі

Для вирішення завдання класифікації обрано модель Random Forest [15]. Random Forest - це метод машинного навчання, що використовується для класифікації, регресії та інших задач. Це ансамблевий метод, де ухвалення рішень відбувається на основі побудови множини дерев. Random Forest має певні особливості, серед них : наявність функції визначення значущості змінних, балансування класів, обробки пустих значень.

Алгоритм працює з p-вимірним випадковим вектором $X = (X_1,..., X_p)$, що представляє предикторні змінні. Тобто, кожне дерево навчається на випадковій вибірці тренувальних даних : у такій вибірці одні і ті ж приклади даних можуть зустрічатись по декілька разів, а деякі - не потраплять до вибірки зовсім. Y - цільова змінна або відповідь. Ціллю роботи Random Forest є знаходження функції передбачення $f(X)$ для прогнозування змінної Y . Функція передбачення визначається функцією втрат $L(Y, f(X))$ для мінімізації очікуваного значення втрати $E_{XY}(L(Y, f(X)))$, тобто $L(Y, f(X))$ визначає, наскільки передбачене значення $f(X)$ відхиляється від реального Y. Для класифікації використовується zero-one loss $L(Y, f(X)) = I(Y \neq f(X))$, що приймає значення 0, якщо $Y = f(X)$, і 1 в інших випадках.

Спочатку усі значення bootstrap-вибірки містяться в одному вузлі. Також для кожного вузла дерева під час поділу обирається випадкова множина ознак, що стає на заваді сильному впливу найбільш інформативних ознак на результат прогнозу усіх дерев. Дерева будуються незалежно один від одного, за допомогою методу бінарного рекурсивного поділу. Кожне дерево виконує послідовний розподіл простору значень вхідних змінних і формує вузли. Кожен вузол-нащадок обробляється аналогічно до батьківського.

Конкретне розбиття, яке дерево використовує для поділу вузла на два вузли-нащадки, обирається шляхом перегляду всіх можливих розбиттів за кожною предикторною змінною та вибору найкращого варіанту відповідно до певного критерію. Для неперервних змінних, дані, значення яких менші за обраний поріг, переходять у лівий вузол, а решта – у правий. Така процедура триває рекурсивно до виконання критерію зупинки. Таким критерієм може бути мінімальна кількість елементів у вузлі або досягнення чистих класів для класифікаційних задач.

Після завершення навчання, для класифікації нових даних, кожне дерево формує свій прогноз - вибирає до якого класу вони відносяться. Далі відбувається голосування між усіма деревами, результатом чого є фінальний прогноз :

$$f(x) = argmax_y \sum_{j=1}^{J} I(h_j(x) = y)$$, де

- $h_j(x)$ - прогноз, зроблений j-тим деревом;
- I - індикаторна функція;
- argmax - означає, що вибирається такий клас y, який отримав найбільшу кількість голосів.

Варто зауважити, що при використані bootstrap-вибірки, можуть бути такі зразки даних, які не увійшли до жодної випадкової вибірки, а отже не були враховані. Такі дані називаються out-of-bag даними. З Out-of-bag даними є можливість працювати як з тестовим набором. У такому випадку немає необхідності розбиття датасета на тренувальний та тестовий набори. Кожне

спостереження перевіряється лише на тих деревах, які не використовували його під час тренування. Це називається out-of-bag-прогнозуванням.

Random Forest використовує незвичайний метод оцінки важливості ознак. Для того, щоб оцінити важливість певної ознаки, спочатку out-of-bag дані проходять через дерево і формується прогноз. Далі значення ознаки випадково переставляються. Модифіковані out-of-bag дані проходять через дерево і формується новий прогноз. Таким чином, маємо дві множини прогнозів : для початкових та модифікованих даних. Для задачі класифікації, важливість ознак визначається як різниця між рівнем помилки отриманих прогнозів. Рівень помилки класифікації обчислюється за формулою :

$$E = \frac{1}{N}\sum_{i=1}^{N} I(y_i \neq f(x_i)), \text{ де}$$

- N- загальна кількість спостережень;
- y_i - справжнє значення класу;
- $f(x_i)$ - прогнозований клас;
- $I(y_i \neq f(x_i))$ - індикатор неправильної класифікації (1, якщо прогноз невірний, 0 — якщо правильний).

Для вирішення проблеми дисбалансу класів було обрано метод SMOTE [16]. Основна ідея методу Synthetic Minority Over-sampling Technique полягає у генерації синтетичних зразків для меншості класу, а не просто дублювання існуючих. Для кожного зразка меншості обирається певна кількість найближчих сусідів. Створюється новий синтетичний зразок на відрізку між вихідним прикладом і одним з сусідніх значень. Нові синтетичні зразки створюються за формулою :

$$X_{new} = X_{original} + \lambda \cdot (X_{neighbor} - X_{original})$$

Алгоритм роботи методу SMOTE :

1. Визначити кількість зразків меншого класу, яку потрібно створити.
2. Для кожного зразка меншості визначити кількість найближчих сусідів.

3. Випадково вибрати одного з сусідів та обчислити різницю між вихідним значенням зразку меншості.
4. Результат помножити на випадкове число, що належить проміжку від 0 до 1.
5. Отримане значення додати до вихідного зразка - створити новий синтетичний зразок.

Створені зразки додаються до початкового набору даних, що сприяє покращенню роботи методів машинного навчання. Метод SMOTE зменшує схильність до перенавчання, що може виникати при звичайному дублюванні зразків. Використання SMOTE допомагає розширити область класу меншості та покращує продуктивність алгоритмів машинного навчання.

2.6 Методика обробки отриманих результатів класифікації

Кібернетичні атаки на об’єкти критичної інфраструктури часто бувають помітними лише на етапі їх фізичного прояву [17, 18]. Дослідження різного роду аномалій може допомогти виявити інциденти та дисфункції в роботі об’єкта та вчасно запобігти негативним наслідкам, які часто можуть стати критичними для навколишнього середовища та суспільства [19].

Після роботи класифікатора дуже важливо правильно обробити отримані дані, які поділяються на дані нормальної роботи системи та аномалії. Для забезпечення кібербезпеки першочергово потрібно визначити причину виникнення аномалії, за допомогою системи прийняття рішень. Наприклад, причиною аномальних даних у показниках систем водопостачання може бути технічний збій у роботі обладнання або вплив кібератаки.

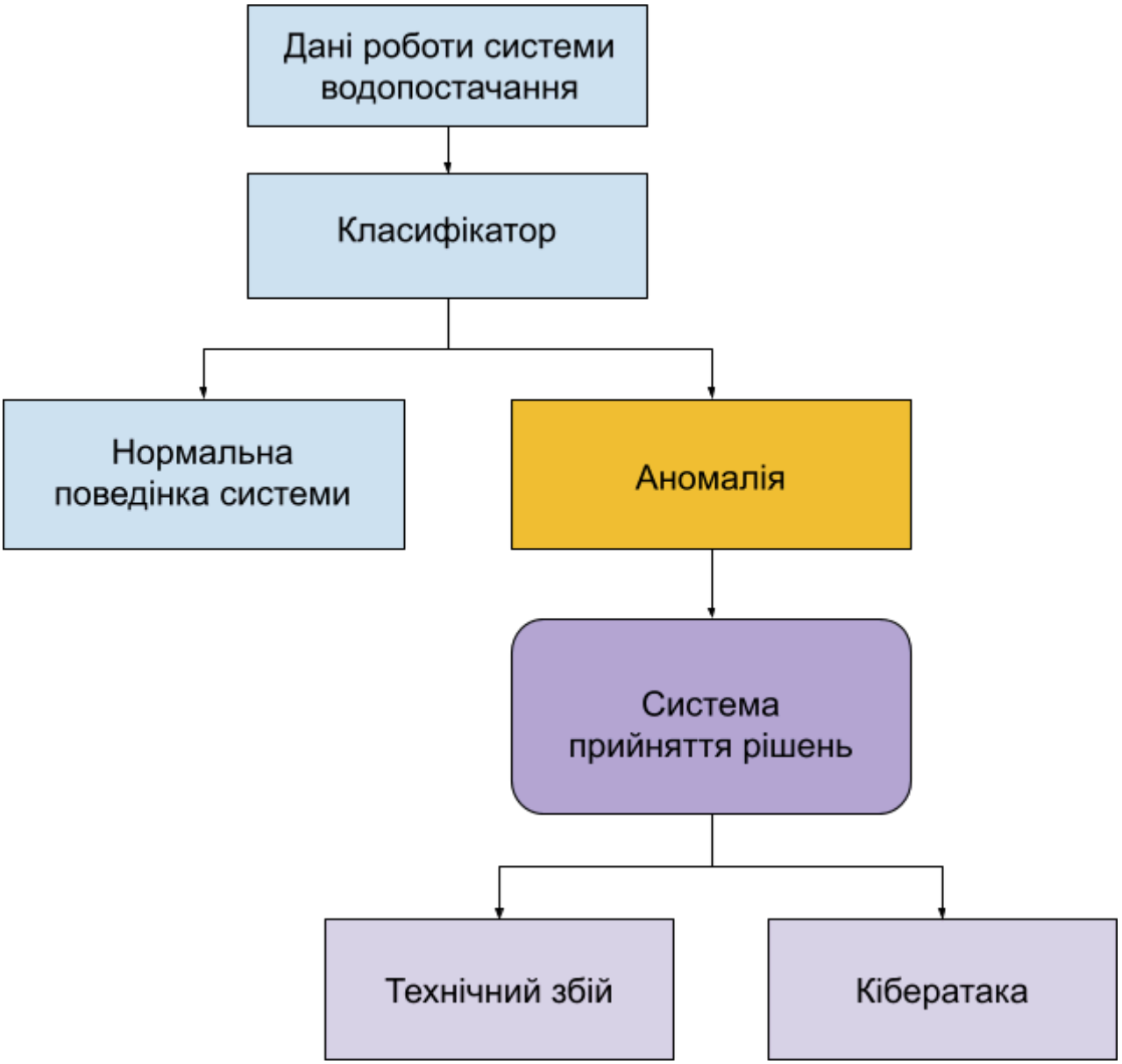

Рисунок 2.9 - Сценарій дій для обробки аномалій

Для визначення чи є аномалія кібератакою або технічним збоєм обладнання, система прийняття рішень може працювати на основі аналізу лог-файлів :

- логи безпеки (журнали подій систем безпеки, таких як : IDS/IPS, антивіруси, брандмауери, системи контролю доступу та інші);
- логи розумних пристроїв (дані програмованих логічних контролерів, стани пристроїв, дані з сенсорів та інформація про взаємодію з користувачем);
- логи мережевих пристроїв (журнали подій, таблиці маршрутизації, доступність шлюзів, мережевий трафік та сесії);
- логи комутаторів (MAC-адреси, журнальні файли портів, стани з’єднань);

- historian server (журнали історичних даних).

У сучасному світі кібератаки є дуже різноманітними. Вони по різному проявляються і потребують правильного реагування. Наприклад, сліди DoS/DDoS атаки можна знайти у логах брандмауера, IDS-системи, мережевих логах та у SCADA-журналах. У такому випадку потрібно перевірити джерела трафіку, обмежити доступ з підозрілих IP, увімкнути фільтрацію трафіку або активувати DDoS-захист, якщо є, і варто тимчасово перевести критичні сервіси на резервні канали.

Сліди FDI-атаки можна знайти у логах historian-сервера, де фіксуються різкі або нетипові зміни показників, у SCADA-журналах, де з’являються неочікувані події або стани обладнання, а також у логах аномалій сенсорів, якщо є розбіжності між фізичними значеннями і цифровими даними. У такому випадку потрібно порівняти дані з кількох джерел, перевірити їх достовірність, застосувати алгоритми виявлення аномалій і, за потреби, провести польову перевірку обладнання.

Сліди шкідливого ПЗ можна знайти у логах антивірусного захисту та EDR-систем, де з’являються сигнали про підозрілу активність, а також у журналах контролерів, якщо відбувалися несанкціоновані зміни, і варто звернути увагу на мережеві логи, якщо спостерігається зв’язок із невідомими або шкідливими доменами. У відповідь необхідно ізолювати уражені системи, провести аналіз логів та пам’яті, а також очистити систему або перевстановити компоненти, та перевірити, чи не поширилася загроза далі по мережі.

Replay-атаки можна виявити у мережевих логах, якщо спостерігається повторення однакових пакетів або затримка в оновленні даних, а також у журналах historian сервера, якщо дані залишаються статичними при змінних умовах середовища. Для реагування потрібно перевірити часові мітки переданих даних, налаштувати механізми виявлення повторюваних записів, застосувати криптографічні методи перевірки автентичності і, за потреби, перевірити дані на об'єкті вручну.

У таблиці представлені найбільш поширені атаки на об’єкти критичної інфраструктури та де вони проявляються.

Таблиця 2.2 - Кібератаки на об’єкти критичної інфраструктури [17]

Тип атаки	Де проявляється
DoS/DDoS	Журнали брандмауера, IDS, маршрутизаторів, SCADA-журнали, MTU
FDI	Журнали historian server, SCADA-журнали подій, сервер БД, RTU, PLC.
Replay	MTU, SCADA, RTU, historian server
Covert	RTU, MTU, SCADA, маршрутизатори
Time delays	MTU, SCADA, PLC, RTU, сервер БД, логи безпеки
Physical attacks	RTU, PLC, сенсори, журнали подій, журнали безпеки
Spoofing	Domain controller, маршрутизатори, робочі станції
Sniffing	Логи безпеки, IDS, маршрутизатори
TSA	SCADA, MTU, RTU, сервер БД
Malware	Робочі станції, domain controller, журнали сервера

Яку саме інформацію з пристроїв та журналів різних систем варто використовувати для виявлення кібератак описано у таблиці 2.3.

Таблиця 2.3 - Важлива інформація для виявлення кібератак [20]

Пристрій	Артефакти	Важлива інформація
Master Terminal Unit	Журнали подій	Містять інформацію про події, які відбуваються в системі, включаючи підключення та відключення пристроїв, передачу даних, виконання команд тощо.
	I/O дані	Це дані введення/виведення з польових пристроїв. Аналіз цих даних може розкрити аномалії, які можуть свідчити про проблеми в системі.

Продовження таблиці 2.3

Пристрій	Артефакти	Важлива інформація
Master Terminal Unit	Конфігураційні дані	Дані про польові пристрої, включаючи параметри зчитування, інтервали опитування тощо. Ці дані можуть бути використані для аналізу структури системи та виявлення змін у конфігурації.
Маршру-тизатори	Журнали подій	Містять інформацію про всі події та дії, що стосуються маршрутизації даних. Це можуть бути записи про встановлення з'єднань, відправлення та отримання пакетів, помилки маршрутизації тощо.
	IP-tables	Містять інформацію про правила маршрутизації, тобто які IP-адреси або діапазони IP-адрес спрямовуються через конкретні інтерфейси маршрутизатора.
	MAC-адреси	Інформація про MAC-адреси пристроїв, які були підключені до маршрутизатора, може бути корисною для виявлення пристроїв у мережі.
	Таблиці маршрутизації	Надають інформацію про шляхи та маршрути для доставки даних між різними мережами або сегментами.
Domain - controller	Події входу в систему	Містять інформацію про всі спроби входу в систему користувачами. Це включає усі входи за допомогою облікових записів домену, а також спроби входу з зовнішніх джерел автентифікації, таких як RADIUS або LDAP.
	Події безпеки	Містять інформацію про різні події, пов'язані з безпекою, такі як спроби неуспішної автентифікації, зміни прав доступу, заборонені спроби доступу тощо.
	Аудиторські події	Містять інформацію про різні події, пов'язані з безпекою, такі як спроби неуспішної автентифікації, зміни прав доступу, заборонені спроби доступу тощо.

Продовження таблиці 2.3

Пристрій	Артефакти	Важлива інформація
Domain - controller	Події зміни паролю	Містять інформацію про всі зміни паролів користувачів в домені. Це дозволяє виявити зміни, які не були здійснені власниками облікових записів або зміни паролів безпеки.
Робочі станції	Виконання програм/файлів	Надає інформацію про те, які програми були запущені на робочій станції і які файли були виконані.
	Використання облікових записів	Це інформація про входи користувачів на робочій станції, зміни паролів, створення або видалення облікових записів тощо.
	Журнали підключених пристроїв	Інформація про підключення зовнішніх пристроїв до робочої станції (наприклад, USB-пристрої, принтери тощо).
Корпора-тивний сервер	Журнали сервера	Містять інформацію про різні події і дії, які відбуваються на робочому сервері (наприклад, запуск/зупинка служб, доступ до систем, виконання конфігураційних змін).
	Журнали подій	Включають інформацію про помилки, збої, зміну файлів або налаштувань системи.
Сервер БД	Файли даних	Ці дані можуть містити інформацію про транзакції, зміни в БД, спроби доступу або інші операції.
	Журнали сервера	Містять інформацію про різні події, що стосуються роботи бази даних, такі як запити, входи користувачів, зміни структури тощо.
	Журнали подій системи	Містять інформацію про різні системні події, які стосуються операційної системи сервера баз даних. Це можуть бути помилки, попередження, запуск служб тощо.
	Трасувальні файли	Зберігають детальну інформацію про виконання запитів до бази даних, включаючи відомості про час виконання, використання ресурсів, операції з блокуванням тощо.

Продовження таблиці 2.3

Пристрій	Артефакти	Важлива інформація
Пошто - вий сервер	Журнали сервера	Містять інформацію про різні події, що стосуються роботи сервера, такі як прийом та відправлення повідомлень, спроби автентифікації користувача, зміни конфігурації тощо.
	Поштові транзакції	Це інформація про всі внутрішні та зовнішні відправлення та отримання електронних листів, включаючи метадані (наприклад, адреси відправника та одержувача, час відправлення та отримання тощо).
Веб сервер	Журнали сервера	Містять інформацію про всі запити, які надходять до сервера, включаючи інформацію про відправників (IP-адреси), URL-адреси, типи запитів (GET, POST тощо), статуси відповідей (успішні або помилкові) та інші важливі дані.
	Журнали подій	Містять інформацію про різні події, які стосуються роботи самого сервера або операційної системи, на якій він працює. Це можуть бути помилки, попередження, запуск або зупинка служб тощо.
	Дані сеансів	Інформація про активні сеанси (наприклад, ідентифікатори сеансів, час початку та завершення сеансів).
Firewall IDS IPS	Дані журналів	Містять інформація про всі мережеві події, що стосуються трафіку, який проходить через пристрій. Ці дані можуть включати інформацію про вхідні та вихідні з’єднання, блоковані або дозволені запити, виявлені атаки тощо.
	IP адреси	Інформація по IP адреси, які взаємодіють з пристроями фаєрвола, IDS, IPS, дозволяє виявляти підозрілу або незвичайну активність, таку як сканування портів, атаки типу “людина посередині” або вторгнення.

Кінець таблиці 2.3

Пристрій	Артефакти	Важлива інформація
Firewall IDS IPS	Журнальні файли портів	Містять інформацію про використання мережевих портів пристроєм, включаючи вхідні та вихідні з'єднання, статуси портів тощо.

Запропонований підхід виявлення аномалій може використовуватись для вирішення важливих задач кібербезпеки :

- оптимізація ознак для моделей безпеки (підвищення ефективності систем виявлення аномалій та виділення критичних системних параметрів для моніторингу);
- виявлення аномалій у роботі систем ;
- підвищення точності IDS/IPS систем (синтетичні дані дозволяють змоделювати рідкісні або нові сценарії атак).

Отримані результати роботи класифікатора можна використовувати як вихідні дані для системи прийняття рішень для визначення причини появи аномалії і своєчасного правильного реагування у випадку кібератаки. Існуючі атаки мають свої власні патерни, що допомагає визначити чи аномалії спричинені кіберфізичною атакою, яка змінює фізичні параметри, чи це фізичний збій у роботі обладнання. Визначення приналежності ситуації до певного паттерна може відбуватись з використанням таких методів, як байєсове виявлення з бінарною гіпотезою, метод зважених найменших квадратів , x^2-детектори на основі фільтрів Калмана і техніки виявлення та ізоляції несправностей [18].

Висновки до Розділу 2

У другому розділі було конкретизовано сферу дослідження. Також було проаналізовано та описано наявні датасети для сфери водопостачання. Після

детального аналізу, обрано датасет для дослідження : The BATtle of the Attack Detection ALgorithms .

Було вивчено та описано роботу стенда системи резервуарів в лабораторії ICS НН ФТІ - ASTANK2. Механізм впливу на фізичні параметри за допомогою кібератаки було досліджено на базі лабораторії ICS, що допомогло визначити загальні характеристики для визначення аномалій. Також було описано ознаки для виявлення аномалій для обраного набору даних BATADAL.

Було обрано метод розв’язання поставленої задачі шляхом класифікації, а також вибрано метод для роботи з дисбалансом класів. Запропоновано використання системи прийняття рішень для визначення причини появи аномалії на основі даних результату роботи класифікатора.

3 ПРАКТИЧНИЙ ЕКСПЕРИМЕНТ

3.1 Підготовка датасету

Для тестування запропонованого підходу для покращення класифікації було обрано тренувальний розмічений датасет BATADAL. Перед початком роботи з датасетом, було виконано його підготовку для покращення роботи класифікатора. За допомогою бібліотеки pandas, було завантажено датасет.

Далі було виконано обробку назв стовпців, а саме : видалено зайві пробіли для уникнення можливих помилок під час подальшої роботи з датасетом. Оскільки датасет є розмічений, то було відокремлено мітки та ознаки (усі колонки в датасеті, крім останньої).

Пропущені значення, якщо такі були, заповнювались середнім значенням для відповідного стовпця. Для забезпечення однакового масштабу ознак, виконано стандартизацію за допомогою класу StandardScaler з бібліотеки scikit-learn.

Оскільки клас міток вихідного датасету містив значення -999, що означало нормальну поведінку системи, та 1, тобто аномалія, для зручності, під час подальшої роботи, було замінено значення міток нормальної поведінки на 0.

Далі було сформовано новий датасет, з вже підготовленими даними ознак і додано стовпець з мітками класів, та збережено його, як новий файл формату csv.

3.2 Оцінка важливості характеристик

Для навчання моделі Random Forest, було відокремлено ознаки та мітки датасета. Оцінювання важливості ознак відбувалось за допомогою вбудованої в методі машинного навчання Random Forest функції оцінювання важливості.

```
X_train = train_df.drop(columns=['label'])
y_train = train_df['label']

rf = RandomForestClassifier(n_estimators=100, random_state=42)
rf.fit(X_train, y_train)

feature_importances = rf.feature_importances_
```

Рисунок 3.1 - Оцінка важливості ознак

Отриманий результат було відсортовано від найважливіших ознак до найменш важливих і побудовано смугову діаграму :

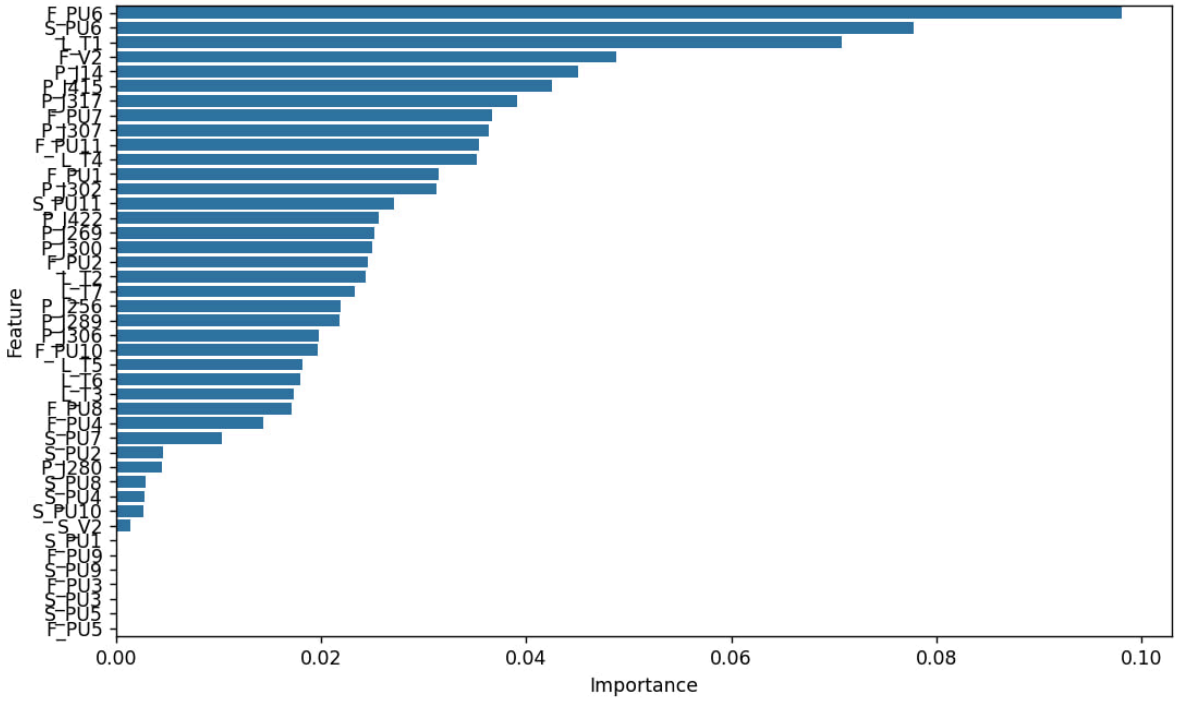

Рисунок 3.2 - Результати оцінки важливості характеристик

За результатами оцінки важливості ознак датасета, можна виділити ряд характеристик, які є неінформативними, а саме : S_PU1, F_PU9, S_PU9, F_PU3, S_PU3, S_PU5, F_PU5. Найбільш інформативними ознаками виявились F_PU6, S_PU6 та L_T1.

3.3 Програмна реалізація моделі класифікації

Для програмної реалізації моделі класифікації Random Forest використовувались python-бібліотеки pandas та scikit-learn. Спочатку завантажуємо необхідні бібліотеки, функції та класи :

- pandas - бібліотека для роботи з даними;
- train_test_split - функція, що використовується для розділення датасету на тренувальну та тестову вибірки;
- RandomForestClassifier - класифікатор на основі ансамблю дерев рішень (Random Forest);
- confusion_matrix - матриця неточностей, що показує, як модель класифікувала приклади з кожного класу (містить значення TN, FN, TP, FP);
- f1_score - функція для оцінювання ефективності моделі класифікації, за метрикою F1 score

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix, f1_score
```

Рисунок 3.3 - Імпорт необхідних бібліотек

Далі завантажуємо датасет у змінну data, за допомогою бібліотеки pandas, та визначаємо набір ознак. Перший набір ознак містить три найбільш значущі ознаки, відповідно до попередньо проведеного ранжування за важливістю, далі набір ознак поступово розширюватиметься.

```
data = pd.read_csv('prep_balanced_dataset.csv')

#перший набір ознак
feature_columns = ['F_PU6', 'S_PU6', 'L_T1']
```

Рисунок 3.4 - Завантаження датасету та визначення набору ознак

Наступним кроком, задаємо набір ознак та цільову змінну. Змінна X містить дані відповідного набору ознак, який заданий змінною feature_columns. Змінна Y є цільовою змінною, яку потрібно передбачити, і містить значення колонки label, тобто мітки класів.

```
X = data[feature_columns]
y = data['label']
```

Рисунок 3.5 - Визначення набору ознак та цільової змінної

За допомогою функції train_test_split розділяємо датасет на тренувальну та тестову вибірки. Дані розділяються за принципом 90/10 :

- 90% даних використовуються для тренування моделі і містяться у змінних X_train, y_train ;
- 10% даних призначені для тестування класифікатора (X_test, y_test).

Параметр stratify = y забезпечує розділення датасету зі збереження пропорції класів у тренувальній та тестовій вибірках.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=42, stratify=y)
```

Рисунок 3.6 - Розділення датасету на тренувальну та тестову вибірки

Після того, як визначені всі необхідні змінні та параметри, будуємо модель класифікатора, яку визначаємо у змінній model, за допомогою RandomForestClassifier. Навчаємо створену модель на тренувальній вибірці, за допомогою функції fit. Модель аналізує дані змінних X_train, y_train та будує внутрішню структуру дерев рішень.

```
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)
```

Рисунок 3.7 - Побудова та навчання моделі

Далі модель робить передбачення цільової змінної для тестової вибірки, використовуючи раніше побудовану структуру дерев рішень.

```
y_pred = model.predict(X_test)
```

Рисунок 3.8 - Передбачення цільової змінної

Для оцінки ефективності роботи моделі класифікації створюємо матрицю неточностей, що містить такі складові :

- TN : істинно-негативне передбачення : аномалії немає і модель також це передбачила ;
- FN : хибно-негативне передбачення : аномалія є, але модель її не передбачила ;
- FP : хибно-позитивне передбачення : аномалії немає, але модель визначила, що це аномалія ;
- TP : істинно-позитивне передбачення : вихідні дані є аномальними і модель також передбачила аномалію.

За допомогою метода .ravel() перетворюємо матрицю в одновимірний масив для зручності використання. А також обчислюємо значення метрики F1 score.

```
conf_matrix = confusion_matrix(y_test, y_pred)
tn, fp, fn, tp = conf_matrix.ravel()

print(f"False Positives (FP): {fp}")
print(f"False Negatives (FN): {fn}")
print(f"F1 Score: {f1_score(y_test, y_pred):.4f}")
```

Рисунок 3.9 - Обчислення та вивід результатів основних метрик

3.4 Практичний експеримент

У межах експерименту було реалізовано декілька варіантів моделі класифікації з використанням алгоритму Random Forest. Для кожного варіанту використовувалися підмножини ознак, сформовані на основі попереднього ранжування за важливістю. Було сформовано шість наборів ознак, а саме :

1. Ознаки 1-3 : ['F_PU6', 'S_PU6', 'L_T1'];

2. Ознаки 1-7 : ['F_PU6', 'S_PU6', 'L_T1', 'F_V2', 'P_J14', 'P_J415', 'P_J317'];
3. Ознаки 1-13 : ['F_PU6', 'S_PU6', 'L_T1', 'F_V2', 'P_J14', 'P_J415', 'P_J317', 'F_PU7', 'P_J307', 'F_PU11', 'L_T4', 'F_PU1', 'P_J302'];
4. Ознаки 1-22 : ['F_PU6', 'S_PU6', 'L_T1', 'F_V2', 'P_J14', 'P_J415', 'P_J317', 'F_PU7', 'P_J307', 'F_PU11', 'L_T4', 'F_PU1', 'P_J302', 'S_PU11', 'P_J422', 'P_J269', 'P_J300', 'F_PU2', 'L_T2', 'L_T7', 'P_J256', 'P_J289'];
5. Ознаки 1-30 : ['F_PU6', 'S_PU6', 'L_T1', 'F_V2', 'P_J14', 'P_J415', 'P_J317', 'F_PU7', 'P_J307', 'F_PU11', 'L_T4', 'F_PU1', 'P_J302', 'S_PU11', 'P_J422', 'P_J269', 'P_J300', 'F_PU2', 'L_T2', 'L_T7', 'P_J256', 'P_J289', 'P_J306', 'F_PU10', 'L_T5', 'L_T6', 'L_T3', 'F_PU8', 'F_PU4', 'S_PU7'];
6. Ознаки 1-36 : ['F_PU6', 'S_PU6', 'L_T1', 'F_V2', 'P_J14', 'P_J415', 'P_J317', 'F_PU7', 'P_J307', 'F_PU11', 'L_T4', 'F_PU1', 'P_J302', 'S_PU11', 'P_J422', 'P_J269', 'P_J300', 'F_PU2', 'L_T2', 'L_T7', 'P_J256', 'P_J289', 'P_J306', 'F_PU10', 'L_T5', 'L_T6', 'L_T3', 'F_PU8', 'F_PU4', 'S_PU7', 'S_PU2', 'P_J280', 'S_PU8', 'S_PU4', 'S_PU10', 'S_V2'];

Для кожного випадку здійснювалося навчання моделі та обчислення ключових метрик, а саме значення False Positive, False Negative, F1 score. Отримані результати роботи класифікатора представлені у таблиці 3.1 :

Таблиця 3.1 - Результати роботи класифікатора для вихідного датасета

Набір ознак	**False Positive**	**False Negative**	**F1 score**
1 - 3	1	10	0.6857
1 - 7	1	9	0.7222
1 - 13	2	8	0.7368
1 - 22	1	7	0.7895
1 - 30	1	8	0.7568
1 - 36	1	8	0.7568

На основі отриманих результатів, були зроблені висновки про дію метода для різних наборів ознак, а також про достатність зразків у обраному датасеті.

Вихідний датасет складається з 3958 записів, які відносяться до класу 0, та лише 219 записів, що відмічені, як аномалія. З метою усунення дисбалансу класів у датасеті, було застосовано метод синтетичного збалансування Synthetic Minority Over-sampling Technique (SMOTE). Після генерації додаткових зразків для менш представлених класів, кількість записів у датасеті збільшилась до 7916 та було усунено дисбаланс класів. Для збалансованого датасету проведено навчання моделі Random Forest аналогічно: із використанням підмножин ознак, сформованих відповідно до ранжування за важливістю.

Отримані результати для збалансованих даних порівнювались із результатами, отриманими до застосування SMOTE, що дозволило проаналізувати вплив збалансованості класів на якість класифікації.

3.5 Оцінка ефективності запропонованого рішення

Порівняння результатів роботи моделі класифікації для початкового та збалансованого датасетів представлено на Рис. 3.10 - 3.12.

На Рис. 3.11 представлено оцінку помилок другого роду (FN), що показує суттєве зменшення їх кількості, після застосування SMOTE, що є основним фактором підвищення точності моделі. Кількість помилок першого роду (FP) для збалансованого датасета є більшою, порівняно з початковими даними, як видно на Рис. 3.10. Однак, враховуючи те, що спостерігається суттєве зменшення помилок FN після застосування SMOTE, можна зробити висновок про підвищення точності моделі, тобто модель працює суворіше, і пропускає менше аномалій. Таким чином, підвищення здатності моделі виявляти позитивні приклади супроводжується збільшенням кількості хибно позитивних

спрацьовувань, але зберігається позитивна динаміка покращення роботи класифікатора з використанням штучно згенерованих, синтетичних даних.

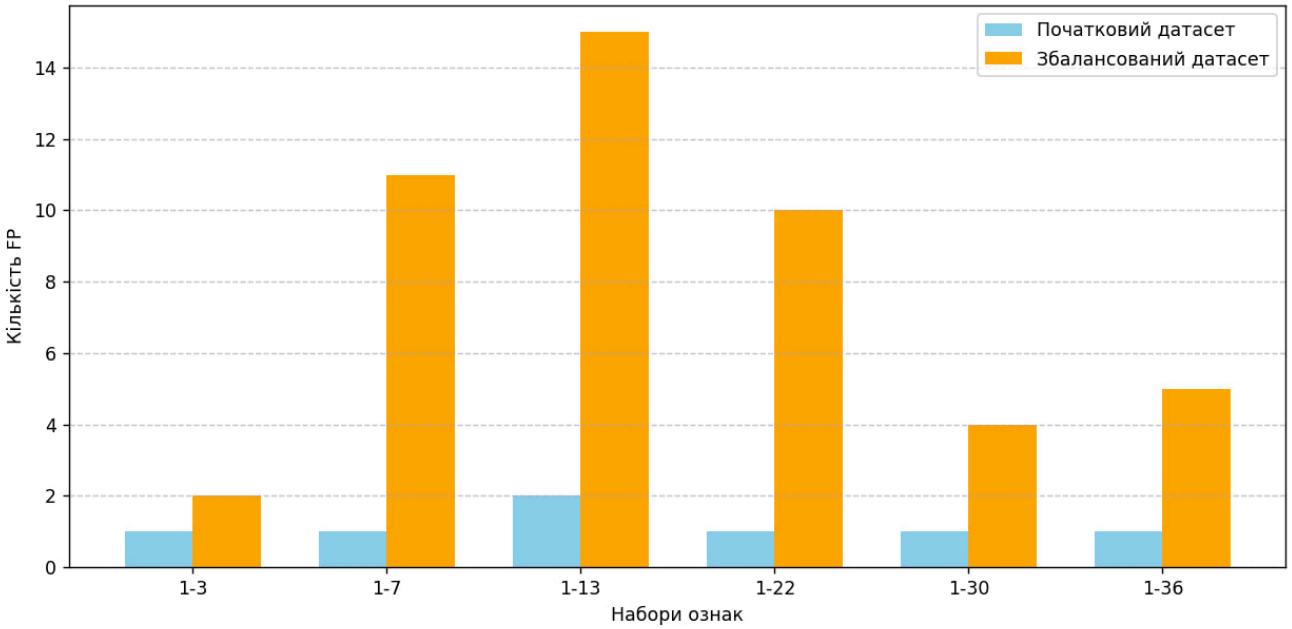

Рисунок 3.10 - Кількість хибнопозитивних помилок для різних наборів ознак початкового та збалансованого датасетів

.

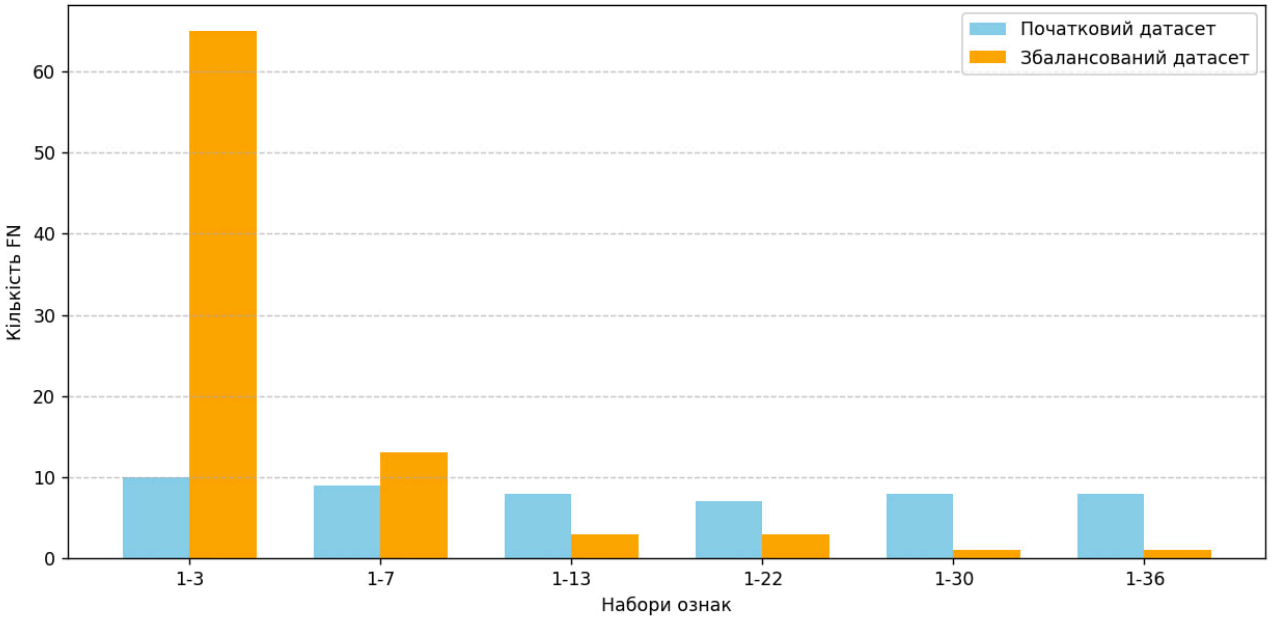

Рисунок 3.11 - Кількість хибнонегативних помилок для різних наборів ознак початкового та збалансованого датасетів

Зі збільшенням кількості ознак спостерігається загальна тенденція до зростання F1 score для початкового та збалансованого датасетів, що показано на Рис.3.12. Також варто зауважити, що результати для 30 та 36 ознак є майже однаковими, тому ознаки 30 - 43 можна вважати неінформативними.

Особливо помітне покращення результатів спостерігається після застосування методу SMOTE. Для збалансованого датасета отримано значно вищі значення F1-міри (до 0.9937) порівняно з початковим (максимум близько 0.78, що є середнім результатом). Це підтверджує важливість усунення дисбалансу класів при роботі з реальними даними, де часто зустрічаються нерівномірні розподіли.

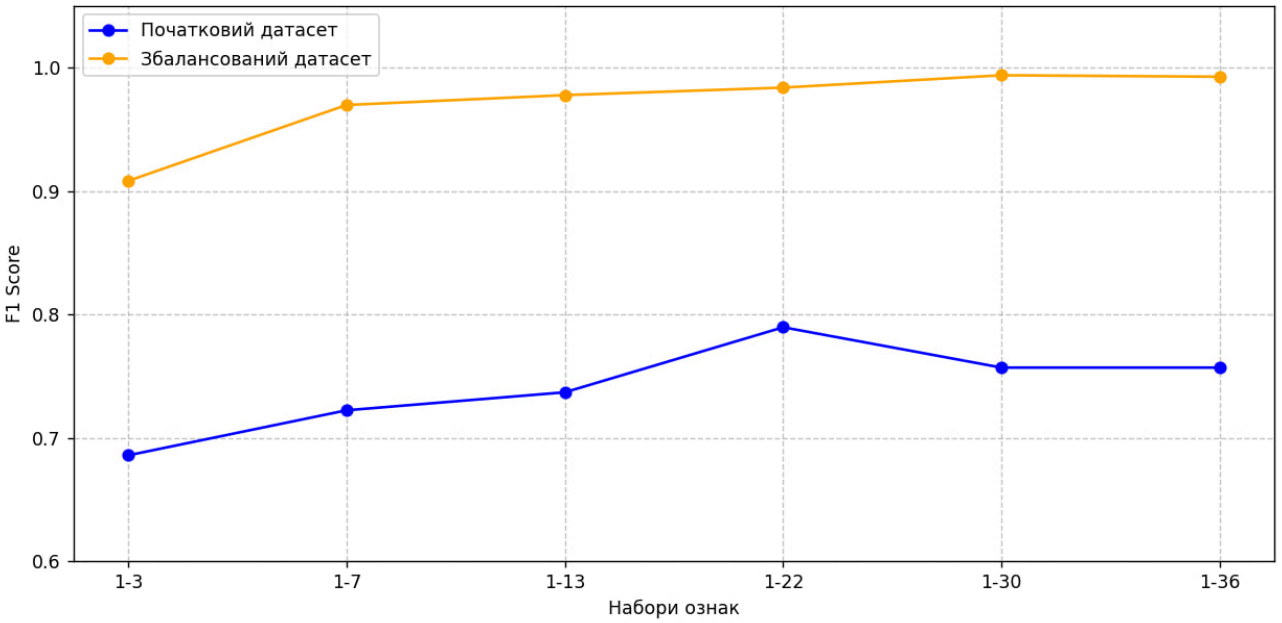

Рисунок 3.12 - Значення F1 score для різних наборів ознак початкового та збалансованого датасетів

За результатами дослідження, можна зробити висновок про суттєвий вплив як кількості ознак, так і збалансованості набору даних на роботу алгоритму Random Forest. При виявленні аномалій у поточній ситуації, необхідно звернути увагу на стан інформаційної системи, а не лише операційної, і проаналізувати стан всіх ключових показників, журналів безпеки. Відхилення фізичних показників є індикатором потенційних атак кіберфізичного типу.

З метою визначення природи аномалії, для датасету BATADAL, відповідно з методикою обробки результатів класифікації, що запропонована у другому розділі, можна здійснити аналіз логів розумних пристроїв (дані програмованих логічних контролерів, стани пристроїв, дані з сенсорів), MTU, маршрутизаторів, контролера домену, робочих станцій, корпоративного сервера, сервера бази даних, поштового та веб серверів, а також логи IDS/IPS чи файєрвола. Такий аналіз допоможе визначити чи аномалія спричинена внаслідок впливу кібератаки та встановити приналежність аномалії до наявних патернів відомих кібератак на об’єкти критичної інфраструктури.

Висновки до Розділу 3

У третьому розділі було описано вибір мови програмування та середовища виконання для реалізації моделі класифікації. Було проведено ознайомлення з особливостями структури обраного датасету. Виконано попередню підготовку датасету для зручності роботи та покращення роботи класифікатора : виконано обробку назв стовпців, оброблено пропущені значення та мітки класів, а також виконано стандартизацію.

З використанням індексу Джинні, було оцінено важливість ознак. Отриманий результат було відсортовано від найбільш значущих до найменш інформативних ознак. На основі ранжування ознак за важливістю було сформовано шість наборів ознак.

У межах експерименту було реалізовано декілька варіантів моделі класифікації з використанням алгоритму Random Forest. Для кожного варіанту використовувалися різні набори ознак. Було оцінено роботу класифікатора на основі отриманих значень ключових метрик : False Positives, False Negatives, F1 score.

Було застосовано техніку синтетичного збалансування Synthetic Minority

Over-sampling Technique (SMOTE). На основі отриманих результатів оцінки роботи класифікатора, експериментально показано, що штучні дані дозволяють покращити якість класифікації. Для збалансованого датасета зі штучними даними отримано значно вищі значення F1 score (до 0.9937) порівняно з початковим (максимум близько 0.78).

Проаналізовано вплив ознак на якість виявлення аномалій : зі збільшенням кількості ознак спостерігається загальна тенденція до покращення якості класифікації. Проведено виділення найбільш значущих ознак, що дозволило розв’язати задачу більш ефективно : найкращі результати оцінки F1 score отримані з використанням набору ознак 1-22 для початкового датасету та 1-30 для збалансованого датасету.

Отримані результати класифікації можуть бути використані як вихідні дані для системи прийняття рішень, що детальніше описано у другому розділі. Існуючі атаки мають свої власні патерни, що допомагає визначити чи аномалії спричинені кіберфізичною атакою, яка змінює фізичні параметри, чи це фізичний збій у роботі обладнання. Перспективою подальших досліджень є встановлення відношення виявлених аномалій до відомих патернів кіберфізичних атак.

ВИСНОВКИ

Під час виконання дипломної роботи, було проаналізовано можливі підходи до детекції аномалій в показних роботи об’єктів критичної інфраструктури з використанням методів машинного навчання . Також обґрунтовано вибір сфери застосування - системи водопостачання.

Аналіз стенда системи резервуарів в лабораторії ICS дав змогу дослідити механізми кібернетичного впливу на фізичні параметри.

В процесі аналізу існуючих датасетів було обрано датасет The BATtle of the Attack Detection ALgorithms. Були описані зміст та основні параметри датасета, система водопостачання C-Town, перелік кібератак, які були зафіксовані обраних наборах даних.

Визначення загального переліку ознак в системах водопостачання дало змогу виділити такі, що свідчать про аномалії та вплив кібератак на фізичні параметри системи, для датасету BATADAL.

У межах практичного експерименту було виконано ранжування ознак за важливістю з використанням індексу Джині у функції оцінки важливості. Було реалізовано декілька варіантів моделі класифікації з використанням методу Random Forest. Також застосовано техніку синтетичного збалансування SMOTE. На основі отриманих результатів оцінки роботи класифікатора, було зроблено висновок про покращення класифікації з використанням штучно згенерованих даних, а також проаналізовано вплив ознак на виявлення аномалій та виділено найбільш значущі набори ознак.

Отримані результати роботи моделі класифікації в подальшому можуть використовуватись, як вихідні дани для системи прийняття рішень для визначення причини появи аномалії та сценарію дій реагування на інцидент. Перспективою подальших досліджень є встановлення відношення виявлених аномалій до відомих патернів кіберфізичних атак.

Практичні результати роботи можуть бути запроваджені до навчального процесу, в тому числі використані в тренінгах на основі лабораторії ICS, а також при адаптації до особливостей об'єкту критичної інфраструктури - використані як додатковий механізм безпеки.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Koay, A. M. Y., Ko, R. K. L., Hettema, H., & Radke, K. (2022). Machine learning in industrial control system (ICS) security: current landscape, opportunities and challenges. Journal of Intelligent Information Systems, 60, 377–405. https://link.springer.com/article/10.1007/s10844-022-00753-1
2. Danyk, Y., Briggs, C., & Maliarchuk, T. (2020). Features of Ensuring Cybersecurity of the Critical Infrastructure of the State. Theoretical and Applied Cybersecurity, 2(1), 84–90 https://www.researchgate.net/publication/343519639_Features_of_Ensuring_Cybersecurity_of_the_Critical_Infrastructure_of_the_State
3. Makrakis, G. M., Kolias, C., Kambourakis, G., Rieger, C. G., & Benjamin, J. (2021). Vulnerabilities and Attacks Against Industrial Control Systems and Critical Infrastructures. arXiv preprint arXiv:2109.03945. https://www.researchgate.net/publication/354493711_Vulnerabilities_and_Attacks_Against_Industrial_Control_Systems_and_Critical_Infrastructures
4. Gawazah, L., Rondla, A., & Balhareth, M. S. A. (2024). To Pay or Not to Pay: The US Colonial Pipeline Ransomware Attack. ResearchGate. https://www.researchgate.net/publication/383206534_To_Pay_or_Not_to_Pay-_The_US_Colonial_Pipeline_Ransomware_Attack
5. Cruceru, A., Wüstrich, L., & Sattler, P. (2022). Review of Industrial Control Systems Protocols. Seminar IITM WS 21/22, Network Architectures and Services, Technical University of Munich. https://www.net.in.tum.de/fileadmin/TUM/NET/NET-2022-07-1/NET-2022-07-1_06.pdf
6. Morris, T. H., & Gao, W. (2013). Classifications of Industrial Control System Cyber Attacks. Proceedings of the 1st International Symposium for ICS & SCADA Cyber Security Research 2013, Leicester, UK

https://www.researchgate.net/publication/311680337_Classifications_of_Industrial_Control_System_Cyber_Attacks

7. Mubarak, S., Habaebi, M. H., Islam, M. R., Abdul Rahman, F. D., & Tahir, M. (2021). Anomaly Detection in ICS Datasets with Machine Learning Algorithms. Computer Systems Science and Engineering, 37(1), 33–46, https://www.researchgate.net/publication/349530048_Anomaly_Detection_in_ICS_Datasets_with_Machine_Learning_Algorithms
8. Culita J., Stefanoiu D., Dumitrascu A. ASTANK2: Analytical Modeling and Simulation // 20th International Conference on Control Systems and Science. — 2015.
9. Dumitrascu, A., , Istratescu, S., Stefanoiu, D., & Culita, J. Environment Communication and Control Systems Integrated on Teaching Platforms // 20th International Conference on Control Systems and Science. — 2015.
10. Secure Water Treatment [Набір даних] / Режим доступу до ресурсу: https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/
11. HIL-based Augmented ICS Security [Набір даних] / Режим доступу до ресурсу: https://github.com/icsdataset/hai
12. Water Storage Tank and Gas Pipeline SCADA systems [Набір даних] / Режим доступу до ресурсу: https://www.impactcybertrust.org/dataset_view?idDataset=1323
13. Water Distribution Testbed [Набір даних] / Режим доступу до ресурсу: https://www.researchgate.net/publication/315849116_WADI_a_water_distribution_testbed_for_research_in_the_design_of_secure_cyber_physical_systems
14. The BATtle of the Attack Detection ALgorithms [Набір даних] / Режим доступу до ресурсу: https://www.batadal.net/data.html
15. Cutler A., Stevens J. R., Cutler D. R. Random Forest // Machine Learning. — 2011. — DOI: 10.1007/978-1-4419-9326-7_5. — URL: https://www.researchgate.net/publication/236952762.
16. SMOTE: Synthetic Minority Over-sampling Technique / N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer // Journal of Artificial Intelligence

Research. — 2022. — P. 321–357. — URL: https://www.researchgate.net/publication/220543125_SMOTE_Synthetic_Minority_Over-sampling_Technique.

17. Cyber Attacks Simulation for Modern Energy Facilities / O. Novikov, M. Shreider, I. Stopochkina, M. Ilin // CEUR Workshop Proceedings. — 2023. — Vol. 3887. — P. 35–49. — URL: https://ceur-ws.org/Vol-3887/ ; Selected Papers of the XXIII International Scientific and Practical Conference "Information Technologies and Security" (ITS 2023).
18. О. Новіков, І. Стьопочкіна, М. Ільїн, М. Овчарук. Визначення параметрів непомітних кібератак на системи керування об’єктів критичної інфраструктури // Наукові вісті КПІ, No 1, с. 69–75, 2024. doi: 10.20535/kpisn.2025.1.322905
19. Oleksii Novikov, Georgy Vedmedenko, Iryna Stopochkina and Mykola Ilin, Cyber Attacks Cascading Effects Simulation for Ukraine Power Grid, p. 23-35//Selected Papers of the XXI International Scientific and Practical Conference "Information Technologies and Security" (ITS 2021), Kyiv, Ukraine, December 9, 2021. URL: http://ceur-ws.org/Vol-3241/
20. Оцінка готовності кіберфізичної системи об’єкту критичної інфраструктури досліджень методами форензики / А. М. Алькова, І. В. Стьопочкіна, О. С. Лиманюк// XXII Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених, Теоретичні і прикладні проблеми фізики, математики та інформатики (13 − 17 травня 2024 р., м. Київ, Україна).- С.243-247
21. Виявлення аномалій як прояву кібернетичних атак на об’єкти водопостачання/ Х. О. Буєва, І. В. Стьопочкіна// XXIII Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених, Теоретичні і прикладні проблеми фізики, математики та інформатики (14 − 17 травня 2025 р., м. Київ, Україна).- С.216-219