

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МІСЬКОГО ГОСПОДАРСТВА ІМЕНІ О. М. БЕКЕТОВА

Пояснювальна записка
до кваліфікаційної роботи бакалавра

на тему:

Аналіз препроцесингу даних для системи керування опаленням будівель

Виконала:

Здобувач вищої освіти, групи ІСтат 2022-1
спеціальності

126 «Інформаційні системи та технології»
(шифр і назва спеціальності)



Емін БАШИРОВ
(прізвище та ініціали)

Керівник:



к.т.н., доц. Володимир БРЕДІХІН
(прізвище та ініціали)

Рецензент:



к.ф-м.н., доц. Олександр КОСТЕНКО
(прізвище та ініціали)

м. Харків – 2026 рік

Харківський національний університет міського господарства імені О. М. Бекетова
(повне найменування закладу вищої освіти)
Навчально-науковий Інститут енергетичної, інформаційної

та транспортної інфраструктури


Кафедра комп'ютерних наук та інформаційних технологій

Рівень вищої освіти перший (бакалаврський)

Спеціальність 126 «Інформаційні системи та технології»
(шифр і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри КНтаІТ


Марина НОВОЖИЛОВА
« 22 » червня 2026 року

З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Баширову Еміну Тофік огли

(прізвище, ім'я, по батькові)

1. Тема роботи Аналіз препроцесингу даних для системи керування опаленням будівель

керівник роботи к.т.н., доцент Бредіхін В.М.

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом закладу вищої освіти від «22» травня 2026 р. № 440-03

2. Термін подання студентом роботи 20.06.2026р.

3. Вихідні дані до роботи Препроцесинг даних для системи керування опаленням будівель





4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

1. Провести аналіз предметної області та існуючих рішень. Дослідити сучасні підходи до побудови систем енергоменеджменту, технології Інтернету речей та методи обробки даних і машинного навчання. 2. Спроектувати архітектуру інформаційної системи. Розробити структурну та функціональну архітектуру системи, що включає підсистеми збору даних із IoT-пристроїв, передачі, зберігання та обробки даних, а також клієнтський інтерфейс. Обґрунтувати вибір технологій і інструментів реалізації. 3. Реалізувати програмну систему та моделі обробки даних. Розробити серверну та клієнтську частини системи, реалізувати модулі збору та попередньої обробки даних, а також інтегрувати алгоритми машинного навчання для аналізу енергоспоживання та виявлення закономірностей. 4. Провести тестування та оцінку ефективності системи.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

Презентація – 15 аркушів

6. Консультанти розділів роботи


| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
|------------|--|----------------|------------------|
| | | завдання видав | завдання прийняв |
| Розділ I | Володимир БРЕДІХІН  | 11.05.2026 | 10.05.2026 |
| Розділ II | Володимир БРЕДІХІН  | 17.05.2026 | 15.05.2026 |
| Розділ III | Володимир БРЕДІХІН  | 21.05.2026 | 30.05.2026 |
| Розділ IV | Вікторія МАЛИШЕВА  | 27.05.2026 | 15.06.2026 |
| | | | |
| | | | |
| | | | |

7. Дата видачі завдання 10.04.2026 р.

КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів кваліфікаційної роботи | Строк виконання етапів роботи | Примітка |
|-------|---|-------------------------------|----------|
| 1 | Вибір теми дипломної роботи | 03.05.2026 | Викон. |
| 2 | Затвердження тем, наукових керівників, завдань та календарного плану підготовки дипломної роботи | 05.05.2026 | Викон. |
| 3 | Написання I розділу | 20.05.2026 | Викон. |
| 4 | Написання II розділу | 25.05.2026 | Викон. |
| 5 | Написання III розділу | 05.06.2026 | Викон. |
| 6 | Написання IV розділу | 10.06.2026 | Викон. |
| 7 | Подання дипломної роботи керівнику | 15.06.2026 | Викон. |
| 8 | Робота по усуненню зауважень керівника, уточнення і доповнення практичного матеріалу, оформлення додатків до роботи | 16.06.2026 | Викон. |
| 9 | Подання доопрацьованого варіанту роботи керівнику | 17.06.2026 | Викон. |
| 10 | Захист матеріалів дипломної роботи на засіданні кафедри | 18.06.2026 | Викон. |
| 11 | Офіційний захист матеріалів дипломної роботи на засіданні екзаменаційної комісії | 22.06.2026 | Викон. |
| | | | |
| | | | |

Студент


(підпис)

Емін БАШИРОВ

(прізвище та ініціали)

Керівник роботи


(підпис)

Володимир БРЕДІХІН

(прізвище та ініціали)

АНОТАЦІЯ

Структура та обсяг роботи. Пояснювальна записка кваліфікаційної роботи здобувача вищої освіти групи ІСтат 2022-1 спеціальності 126 «Інформаційні системи та технології» Баширова Еміна Тофік огли за темою «Аналіз препроцесингу даних для системи керування опаленням будівель» складається з 4 розділів, містить 56 сторінок тексту, 27 рисунків, 5 таблиць, 41 джерело.

Дана робота присвячена актуальній темі розробки інформаційної системи для аналізу та обробки даних у сфері енергоспоживання з використанням сучасних методів машинного навчання та технологій Інтернету речей. Актуальність обумовлена зростанням обсягів даних, що надходять із сенсорних пристроїв, а також необхідністю підвищення енергоефективності та автоматизації процесів управління системами опалення.

Метою дослідження є розробка програмної системи для збору, обробки та інтелектуального аналізу даних енергоспоживання з метою підвищення ефективності функціонування систем опалення.

Об'єктом дослідження є процес моніторингу та аналізу параметрів енергоспоживання в системах опалення.

Предметом дослідження є методи обробки даних, алгоритми машинного навчання та програмні засоби реалізації інформаційної системи для аналізу енергоспоживання.

У першому розділі роботи виконано постановку задачі, проведено аналіз предметної області систем енергоменеджменту та огляд сучасних підходів до обробки даних і застосування машинного навчання. Визначено функціональні та нефункціональні вимоги до системи, а також сформовано основні сценарії її використання.

У другому розділі наведено опис архітектури інформаційної системи, яка включає підсистеми збору даних із IoT-пристроїв, серверну частину для

обробки та зберігання даних, а також клієнтський інтерфейс. Обґрунтовано вибір технологій та інструментів розробки.

У третьому розділі описано реалізацію основних компонентів системи, включаючи модулі обробки даних та моделі машинного навчання для аналізу енергоспоживання. Наведено результати модульного, інтеграційного та функціонального тестування, що підтверджують працездатність і ефективність розробленої системи.

У четвертому розділі розглядаються питання охорони праці.

Ключові слова: МОБІЛЬНИЙ ДОДАТОК, АВТОРИЗАЦІЯ, КОРИСТУВАЧ, БАЗА ДАНИХ, ANDROID ДОДАТОК

ЗМІСТ

| | |
|--|----|
| ВСТУП..... | 7 |
| РОЗДІЛ 1 ЗАГАЛЬНІ ПОЛОЖЕННЯ | 9 |
| 1.1 Постановка проблеми..... | 9 |
| 1.1.1 Аналіз та постановка проблеми препроцесінгу даних | 11 |
| 1.2 Огляд SCADA-системи ПоліТЕР | 12 |
| 1.3 Огляд існуючих рішень..... | 15 |
| РОЗДІЛ 2 ІНФОРМАЦІЙНЕ ТА МАТЕМАТИЧНЕ ЗАБЕЗПЕЧЕННЯ..... | 21 |
| 2.1 Загальна архітектура модуля очищення даних | 21 |
| 2.2. Опис Препроцесора | 22 |
| 2.3 Функціональні та нефункціональні вимоги | 25 |
| 2.4 Варіанти використання системи..... | 26 |
| Висновки до другого розділу | 29 |
| РОЗДІЛ 3 ПРОГРАМНЕ ТА ТЕХНІЧНЕ ЗАБЕЗПЕЧЕННЯ | 31 |
| 3.1 Програмні засоби реалізації | 31 |
| 3.2 Реалізація компонентів препроцесора | 31 |
| 3.3 Опис модульного тестування препроцесора | 36 |
| 3.4.1 Набор даних | 37 |
| 3.4.2 Проведення експериментів для алгоритму відновлення..... | 37 |
| 3.5 Проведення експериментів для алгоритму пошуку викидів..... | 40 |
| Висновок до третього розділу | 42 |
| РОЗДІЛ 4 ОХОРОНА ПРАЦІ..... | 43 |
| 4.1 Організаційно-правові основи забезпечення безпеки праці | 43 |
| 4.2 Характеристика об'єкта та виявлення потенційних небезпек | 44 |
| 4.3 Дослідження ризику реалізації потенційних небезпек на об'єкті проектування та розробка заходів щодо їх попередження | 46 |
| Висновок до четвертого розділу | 51 |
| ЗАГАЛЬНІ ВИСНОВКИ..... | 52 |
| ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ | 53 |
| ДОДАТОК А | |

ВСТУП

Усі сучасні інтелектуальні системи керування опаленням будівель є комплексом датчиків (наприклад: тиску, температури, витрати), систем моніторингу та керування.

Йдеться про Технологію Інтернету речей (IoT, Internet of Things), яка набирає велику популярність на даному етапі розвитку, оскільки здатна максимально автоматизувати роботу із забезпечення комфортної життєдіяльності людини. Додатково такі системи дозволяє отримувати великі масиви даних про різні параметри, що впливають на загальні умови будівлі. Всебічний аналіз масивів даних IoT систем дозволяють своєчасно виявляти випадки значень продуктивності, що відхиляються, і показників енергоефективності будівель за допустимі межі. При розгляді всієї сукупності даних у єдиній системі збільшення кількості джерел може призвести до виконання некоректного аналізу даних у разі збою навіть одного джерела даних [1].

Причиною подібного може бути наступне: знос обладнання, неправильні монтаж та експлуатація обладнання (людський фактор), конструктивні недоліки програмного та апаратного забезпечення, вихід умов експлуатації за межі допустимих значень, а також багато інших подібних факторів. Для забезпечення ефективної та правильної роботи системи вся некоректна інформація повинна виправлятися [2].

У цій роботі буде описано препроцесор, основними функціями якого є вилучення необроблених даних із сховища даних системи та їх обробка [3].

Вона повинна включати: виявлення аномальних значень та їх очищення; відновлення пропущених даних у часовому ряді. Тимчасовий ряд повинен піддаватися мінімальним змінам, оскільки препроцесор спотворює дані, отримані від датчиків. Але це зроблено з тим розрахунком, щоб не вносити в наступний етап системи свідомо некоректних даних, що виникають через різні

збої. Неточність, що вноситься відновленими даними нижче за ту, яку вносять збої.

Метою дослідження є розробка програмної системи для збору, обробки та інтелектуального аналізу даних енергоспоживання з метою підвищення ефективності функціонування систем опалення.

Для досягнення поставленої мети необхідно вирішити такі завдання:

- 1) провести аналіз предметної області та існуючих рішень;
- 2) спроектувати архітектуру інформаційної системи;
- 3) реалізувати програмну систему та моделі обробки даних;
- 4) провести тестування та оцінку ефективності системи.

РОЗДІЛ 1 ЗАГАЛЬНІ ПОЛОЖЕННЯ

1.1 Постановка проблеми

Сучасні системи керування опаленням будівель є складними кіберфізичними системами, що поєднують апаратне забезпечення (датчики, контролери, виконавчі механізми) з програмним забезпеченням, орієнтованим на оптимізацію споживання енергії та підтримку комфортних мікрокліматичних умов. Еволюція таких систем відбувалася від традиційних ручних регуляторів до автоматизованих систем управління з інтелектуальними алгоритмами.

На початковому етапі системи керування ґрунтувалися на простих термостатах, які підтримували задану температуру в приміщенні без врахування зовнішніх умов або енергетичних тарифів. З розвитком мікропроцесорних технологій у 1990–2000-х роках з'явилися програмовані контролери, здатні враховувати добові режими, погодні зміни та графіки використання приміщень.

Сучасний етап характеризується впровадженням інтелектуальних систем керування (Smart Heating Systems), які використовують методи машинного навчання, предиктивне моделювання та елементи Інтернету речей (IoT), рис. 1.1. Такі системи здатні аналізувати великі обсяги даних у реальному часі, адаптувати режими опалення до змін зовнішнього середовища, виявляти неефективності в роботі обладнання та мінімізувати енергоспоживання без втрати комфорту.

Інтеграція систем опалення в розумні будівлі (Smart Buildings), рис. 1.2, сприяє появі нових підходів до управління, коли рішення приймаються на основі аналітичних моделей, що враховують прогноз погоди, теплові характеристики будівлі, історичні дані про споживання енергії, а також поведінку користувачів.



Рисунок 1.1 - Розумна система для керування опаленням [4]

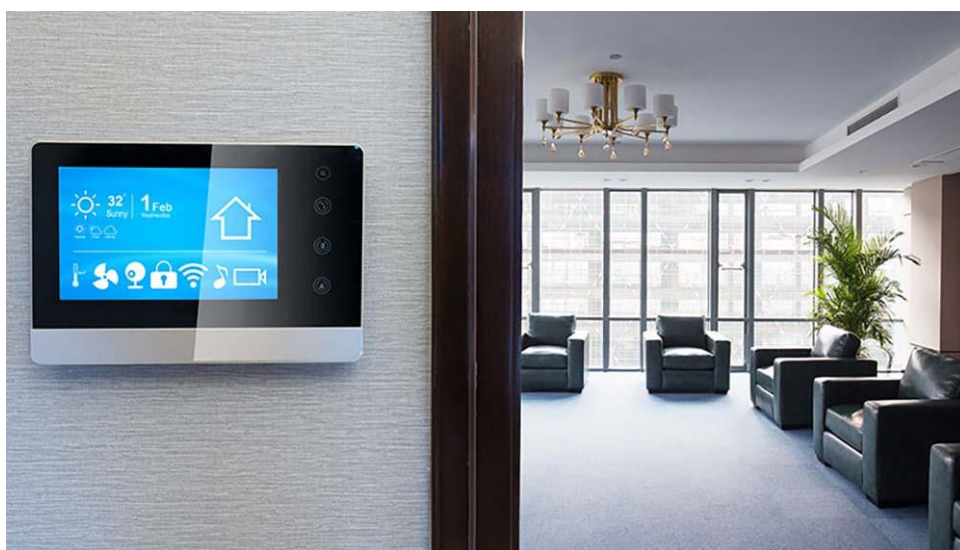


Рисунок 1.2 – Система «розумний будинок» [5]

Таким чином, розвиток систем керування опаленням поступово переходить від реактивного керування до прогнозного та самонавчального, що вимагає ефективної роботи з даними на всіх етапах — від збору до аналітичної обробки.

1.1.1 Аналіз та постановка проблеми препроцесінгу даних

Функціонування сучасних інтелектуальних систем керування опаленням значною мірою залежить від якості даних, що надходять із сенсорів і зовнішніх джерел. Дані можуть надходити з різною частотою, мати пропуски, шум або помилки вимірювання. Тому одним із ключових етапів у побудові моделей оптимального керування є препроцесінг даних — попередня обробка, спрямована на підвищення достовірності, повноти та структурованості інформації, рис. 1.3.

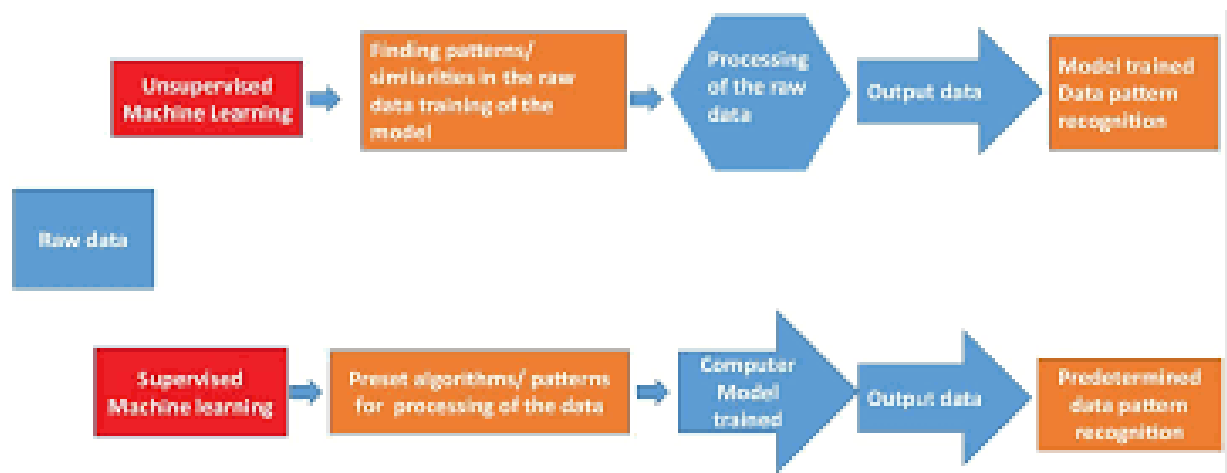


Рисунок 1.3 – Послідовність кроків препроцесінгу даних [6]

Основними проблемами препроцесінгу є:

Шумність і похибки вимірювань. Датчики температури, вологості, тиску та витрати тепла часто мають систематичні або випадкові похибки. Вплив електричних перешкод або неправильне калібрування може спричинити спотворення даних, що унеможливорює точне моделювання процесу теплопередачі.

Пропущені або неповні дані. Через перебої зв'язку або технічні несправності можуть виникати часові прогалини у записах. Їх заповнення потребує застосування методів інтерполяції або відновлення на основі кореляційних зв'язків між змінними.

Нестандартизованість форматів даних. У системах із різними типами сенсорів або протоколів (Modbus, BACnet, KNX) [7] дані можуть мати різну структуру, одиниці виміру або часові інтервали, що ускладнює інтеграцію в єдину базу для аналітики.

Проблема часової синхронізації. Для моделей прогнозування температури або енергоспоживання необхідно синхронізувати показники з різних джерел, включаючи дані погоди, тарифи, графіки використання приміщень тощо.

Виявлення та обробка аномалій. Аномальні значення можуть свідчити про несправність сенсорів або про реальні нештатні режими. Автоматизоване виявлення таких даних є критично важливим для стабільності системи.

Ефективне вирішення цих проблем вимагає створення модуля інтелектуального препроцесінгу, який поєднує методи статистичного аналізу, машинного навчання (наприклад, класифікацію аномалій, фільтрацію даних), а також алгоритми часової обробки сигналів.

Таким чином, препроцесінг даних у системах керування опаленням виступає основою для побудови точних аналітичних і прогнозних моделей, що забезпечують енергоефективність і надійність роботи системи. Недостатня увага до цього етапу призводить до зниження точності керування, перевитрат енергії та підвищення експлуатаційних витрат.

1.2 Огляд SCADA-системи ПоліТЕР

На рис. 1.4 представлена структура інтелектуальної системи управління тепlopостачанням будівлі. Система складається із трьох рівнів. На нижньому рівні система включає різні проводові і бездротові датчики, прилади обліку і контролери. На середньому рівні виконується забезпечення зв'язку контролерів та вузлів обліку з сервером баз даних у вигляді різного дротового та бездротового мережного обладнання. Третій рівень включає SCADA-систему «ПоліТЕР» із сервером баз даних, що здійснює обробку інформації.

Подання оброблених даних здійснюється на АРМ локальних користувачів або на АРМ віддалених користувачів через мережу Інтернет [8].

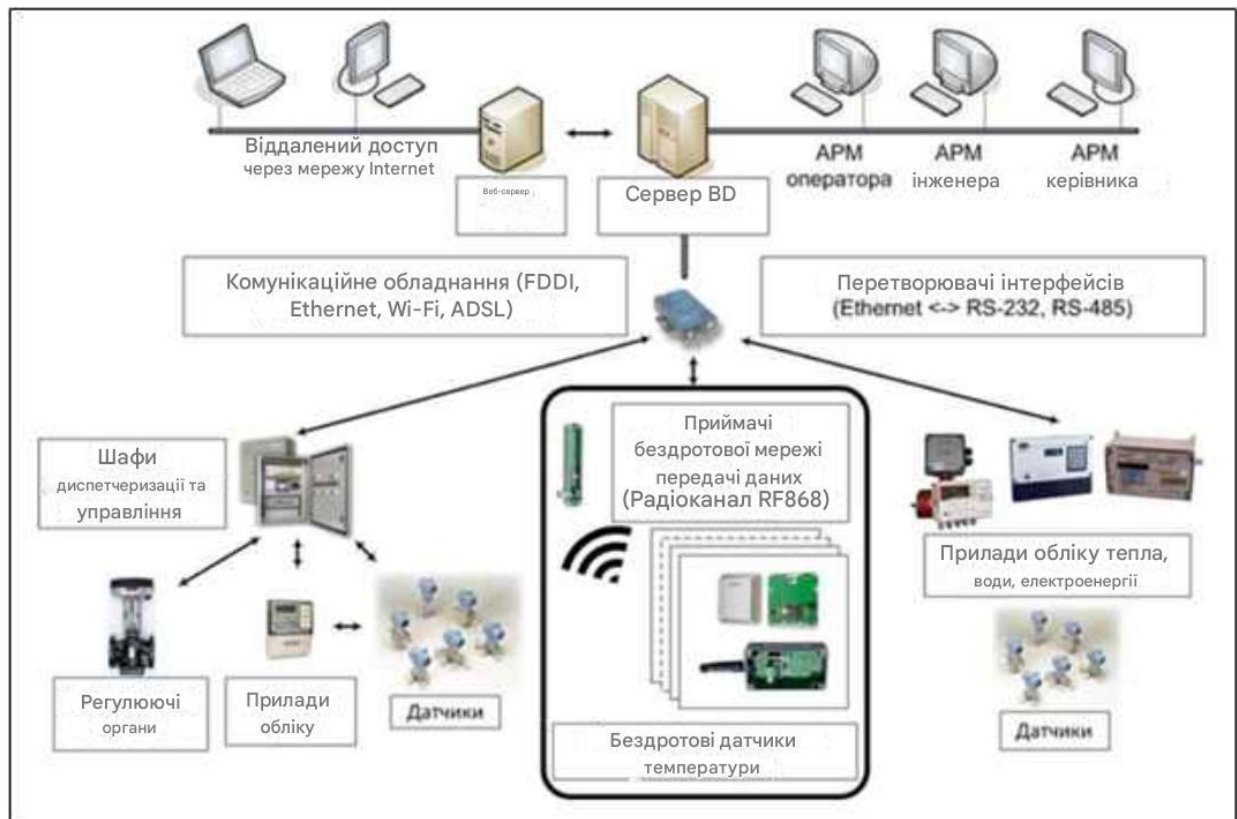


Рисунок 1.4 – Архітектура SCADA-системи ПоліТЕР [9]

Подібні системи використовуються для комплексної оптимізації процесів управління тепlopостачанням та теплоспоживанням за рахунок інтеграції територіально розподіленого вимірювального обладнання з автоматичними системами управління власними системами.

Інтелектуальна система управління тепlopостачанням будівель включає наступне вимірювальне та керуюче обладнання:

- вузли обліку теплової енергії, обладнані теплообчислювачем з підключеними до нього датчиками тиску, датчиками температури та витратомірами;
- вузли обліку холодної води, обладнані обчислювачем з підключеними до нього датчиком тиску та витратоміром;
- вузли обліку газу, обладнані газовим коректором з підключеними до

нього витратоміром датчиками температури та тиску;

- контролери управління розподілом теплової енергії в індивідуальних теплових пунктах з підключеними до них 5 датчиками температури та різними дискретними датчиками захисту, а також керуючими приводами;

- контролери управління генерацією та розподілом теплової та електричної енергії когенераційних установок, газової котельні та центральних теплових пунктів з безліччю різних датчиків та керуючих пристроїв;

- бездротові датчики контролю температурного режиму приміщень.

Програмний комплекс ПоліТЕР реалізований мовами програмування C++, R і використовує Oracle як базову СУБД [10].

Підтримуються відкриті та пропрієтарні протоколи обміну даними з обладнанням різних виробників, конфігуроване середовище візуалізації даних, оповіщення оператора системи SMS та електронною поштою.

Основною проблемою експлуатації вимірювального обладнання є періодична поява помилкових показань чи даних з наявністю обурень невідомого характеру. Однією з причин є вихід ладу: через старіння або порушення умов експлуатації обладнання щорічно 10-20 різних пристроїв підлягають заміні. Іншою причиною є засмічення областей контакту датчика з вимірювальним середовищем. Також періодично виникають обриви ліній зв'язку з контрольно-вимірювальними пристроями, більшість приладів, особливо, контролери, не підтримують апаратні архіви даних. Типовою причиною також є некоректне встановлення обладнання, що призводить до виникнення додаткових збурень, пов'язаних з нетиповими режимами експлуатації IoT-датчиків в області їх встановлення.

Некоректні дані призводять до наступних основних проблем:

- помилкові обчислення під час виконання оптимізаційних алгоритмів, побудови некоректних характеристик об'єктів управління та, як наслідок, до некоректного прийняття стратегічних рішень;

- некоректне ухвалення рішень персоналом за наявності

недостовірних відхилень параметрів експлуатації від їх номінальних значень;

- некоректне автоматичне управління за допомогою контролерів може призвести не тільки до перевитрати енергії, а й до виходу з експлуатації інженерних комунікацій (наприклад, заморожування системи опалення);

- некоректний розрахунок споживаних енергоресурсів або накладення енергопостачальними організаціями на споживача штрафних санкцій за несвоєчасне виявлення та усунення несправностей [11].

Своєчасне виявлення, усунення несправностей обладнання та відновлення даних дає суттєвий організаційно-економічний ефект за рахунок зменшення або виключення втраченої вигоди у разі виникнення зазначених проблем.

В цьому дослідженні запропоновано модуль очищення даних, що розширює систему ПоліТЕР і вирішує наступні основні завдання:

- оперативне виявлення аномальної поведінки датчиків температури та повідомлення оператора про знайдені аномалії;
- оперативне виявлення перепусток і викидів у вимірюваних даних датчиків та заміна таких значень на синтетичні правдоподібні дані.

Модуль очищення даних повинен розташовуватися в системі ПоліТЕР між рівнем зчитування даних та рівнем їх використання для аналітичних розрахунків та подання. При цьому дані, що відображаються, позначаються як вихідні або відновлені, а доступ до некоректних даних зберігається для можливості їх більш глибокого аналізу.

1.3 Огляд існуючих рішень

Виявленню викидів присвячено багато методик для виявлення відмов обладнання [12]. До базових підходів належать методи, що визначають вихід за межі допустимого діапазону відхилення фактичних значень від значень, розрахованих за моделлю об'єкта [13].

У роботі [14] на прикладі систем моніторингу вентиляції та

кондиціювання для визначення некоректних показань використовується поєднання трьох способів обробки даних: зменшення розмірності даних, що корелює, сигналізують про виникнення викидів; застосування розширеного фільтра Калмана для фільтрації шумів та розкладання даних у часові ряди, рис. 1.5; поділ областей коректних та некоректних значень статичних та динамічних параметрів моделі об'єкта з використанням рекурсивного однокласового методу опорних векторів.



Рисунок 1.5 - Застосування розширеного фільтра Калмана для фільтрації шумів

Діяльність [15] розглядається завдання прогностного управління теплопостачанням споживачів з допомогою нейромережі, навченої на статистичних даних, одержуваних з підсистеми моніторингу. Штучна нейронна мережа розглядається як основний інструмент, що мінімізує помилки, пов'язані з ручним керуванням температурою на виході з котельні.

Дослідження підходів до збору та аналізу даних на підприємствах водопостачання та водовідведення розглянуті в роботі [16].

Для визначення відхилень у роботі систем централізованого теплопостачання у роботі [17] пропонується застосовувати метод

найближчого сусіда. У цій роботі як зразок беруться параметри роботи групи подібних автоматизованих індивідуальних теплових пунктів (АІТП). Несправністю вважаються відхилення показників роботи одного АІТП від показників групи схожих з ним АІТП, підключених до подібної системи централізованого теплостачання, на величину заданого порога.

У роботі [18] представлено застосування методу головних компонент для ідентифікації викидів у даних роботи систем опалення, вентиляції та кондиціонування. У роботі [19] описаний спосіб відновлення даних, застосовуваний разом із способом основних компонент.

Також варто відзначити спосіб побудови балансових моделей, що застосовується до пов'язаних теплоенергетичних комплексів. Розбіжність енергетичного чи гідравлічного балансу у таких моделях свідчить про наявність несправностей у системі чи наявність викидів.

Відновлення даних, втрачених у разі викидів, виконується шляхом їх розрахунку з урахуванням методів прогнозування. Безліч робіт присвячено прогнозуванню параметрів систем теплостачання з використанням статичних та динамічних факторних моделей та моделей часових рядів з ідентифікацією методами авторегресійного аналізу [20], методу опорних векторів та рекурентних нейронних мереж.

Також інтерес представляє підхід до відновлення статичних і динамічних характеристик моделей, спотворення яких виникає в результаті періодичних збурень. Облік подібних збурень виконується шляхом опису їхнього часу дії індикаторними функціями, рис. 1.6 [21].

Ідея методів відновлення даних полягає у розрахунку та підстановці замість пропущеного середнього значення до та після пропуску або присвоєння пропущеному – значення попереднього. Але за такого підходу, при великих проміжках відсутніх значень, відновлення не матиме достатньої точності, оскільки вся ділянка буде замінено прямою лінією.

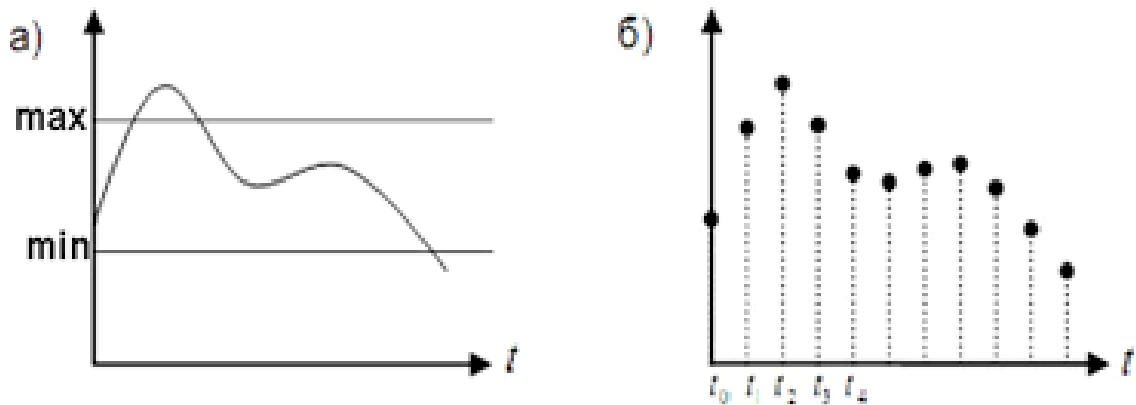


Рисунок 1.6 – Індикаторні функції для відновлення статичних і динамічних характеристик моделей

У роботі [22] розглядаються методи відновлення даних тимчасових рядів, засновані на нейронних мережах, які вирішували проблему відновлення великих проміжків. Автори в роботі [23] представляють метод обчислення в контексті алгоритмів машинного навчання та представляють метод обчислення k -найближчого сусіда (kNNI) для відновлення пропущених значень. Для об'єкта з кількома атрибутами, у якого немає значення для одного атрибута, підхід kNNI шукає k об'єктів з аналогічними значеннями для інших атрибутів відповідно до метрики відстані.

Модель ARIMA [24] – це популярна модель прогнозування часових рядів, яка є узагальненням моделі авторегресії (AR). ARIMA використовує структурні методи аналізу часового ряду, орієнтовані оцінку параметрів майбутніх значень від відомих минулих значень часового ряду при додаткових припущеннях про властивості спостережуваних величин.

Автори роботи [25] відновлення блоків пропущених значень пропонують алгоритм REBOM для нерегулярних часових рядів. Алгоритм будує матрицю, в якій зберігаються тимчасовий ряд з пропущеними значеннями і найбільш лінійно корелюваних тимчасових рядів відповідно до кореляції Пірсона. Вони представляють рішення з лінійною просторовою складністю, засноване на центроїдній декомпозиції (CD), яке є наближенням SVD. Цей алгоритм припускають лінійну кореляцію між підпоследовністю

часового ряду та опорних часових рядів. Якщо часові ряди не корелюються лінійно, точність відновлення погіршується.

ДинаММО використовується для аналізу, класифікації та розрахунку часових рядів, взятих із систем захоплення рухів людини. Він заснований на фільтрах Калмана, які, як і SVD, припускають лінійну кореляцію між часовими рядами для точної оцінки невідомих значень. Алгоритм допускає використання лише одного опорного ряду, цього часто недостатньо для точного відновлення.

MUSCLES [26] – це онлайн-алгоритм, заснований на багатовимірній моделі авторегресії, параметри якої поступово оновлюються за допомогою методу рекурсивних найменших квадратів. Крім минулих значень неповних тимчасових рядів, MUSCLES також враховує найостанніші значення спільно розвиваються і лінійно корелюваних тимчасових рядів, які знаходяться в межах ковзного вікна. Оскільки невеликі неточності відновлення накопичуються протягом тривалого пропущених значень, точність MUSCLES погіршується.

ESPIRIT [27] – алгоритм, що використовує онлайн-аналіз головних компонентів (PCA), щоб скоротити набір з n спільно розвиваються і корельованих потоків до невеликого числа k прихованих змінних, які підсумовують найважливіші тенденції вихідних даних. Для кожної прихованої змінної SPIRIT відповідає одній моделі AR для попередніх значень, яка поступово оновлюється в міру надходження нових даних. Якщо значення немає, моделі AR використовуються для прогнозування поточного значення кожної змінної, з якого виводиться оцінка відсутнього значення. Оновлення моделей з використанням відновлених даних викликає ті ж проблеми, що й MUSCLES, оскільки поширюються неточності. Оскільки PCA і SVD засновані на тому самому базовому принципі, слабкі сторони PCA збігаються з SVD для відновлених часових рядів.

При цьому необхідно зазначити, що особливістю експлуатаційних даних систем теплопостачання є періодичний характер, пов'язаний із циклічною

зміни погодних умов та режимів експлуатації приміщень залежно від часу.

Викиди в експлуатаційних даних, що виникають при роботі різних підсистем, пов'язаних з теплопостачанням, характеризуються різкою зміною сигналу, що вимірюється, що триває в межах обмеженого періоду часу.

У повній системі було вирішено використовувати поєднання кількох підсистем: препроцесор (про який йтиметься далі); предиктор; реконструктор.

У цій роботі піде опис Препроцесора, який включає: пошук і видалення короткочасних аномалій і відновлення даних, які були втрачені датчиками або в результаті роботи очищення від викидів (видалення короткочасних аномалій). Це може забезпечити дуже добрий результат подальшої роботи системи. Так як при короткострокових збоях підготувати якісну навчальну вибірку є одним із першочергових завдань. Нейронні мережі, що навчаються на засмічених даних, можуть показати неякісний вихідний результат, що при прогнозуванні може зробити більше неточності в роботу. Це можуть виправити синтетичні вбудовані в ряд значення після алгоритмів відновлення часового ряду та видалення викидів, представлені далі в цій роботі.

Висновки за розділом

У першому розділі проведено комплексний аналіз предметної області систем керування опаленням будівель. Визначено, що сучасні системи базуються на технологіях IoT та машинного навчання і потребують високоякісних даних для ефективного функціонування. Виявлено ключові проблеми обробки даних, зокрема наявність шумів, пропусків, аномалій та неузгодженості форматів. Проаналізовано існуючі рішення та методи препроцесінгу, що дозволило обґрунтувати необхідність розробки спеціалізованого модуля очищення даних як критично важливого компонента системи.

РОЗДІЛ 2 ІНФОРМАЦІЙНЕ ТА МАТЕМАТИЧНЕ ЗАБЕЗПЕЧЕННЯ

2.1 Загальна архітектура модуля очищення даних

Загальна архітектура Модуля очищення даних представлена на рис. 2.1. Модуль тиражується для кожного температурного датчика системи ПоліТЕР і складається з наступних основних підсистем:

- препроцесор;
- предиктор;
- реконструктор;
- детектор аномалій.

Препроцесор готує накопичені показання датчиків подальшої обробки. Предиктор забезпечує штучну нейронну мережу, яка виконує прогноз наступного значення датчика з урахуванням його історичних даних. Реконструктор визначає, чи є поточне показання датчика викидом, і в цьому випадку замінює його синтетичне значення, отримане Предиктором. Детектор аномалій виявляє аномальні послідовності у показаннях датчика для подальшого повідомлення оператора системи ПоліТЕР.

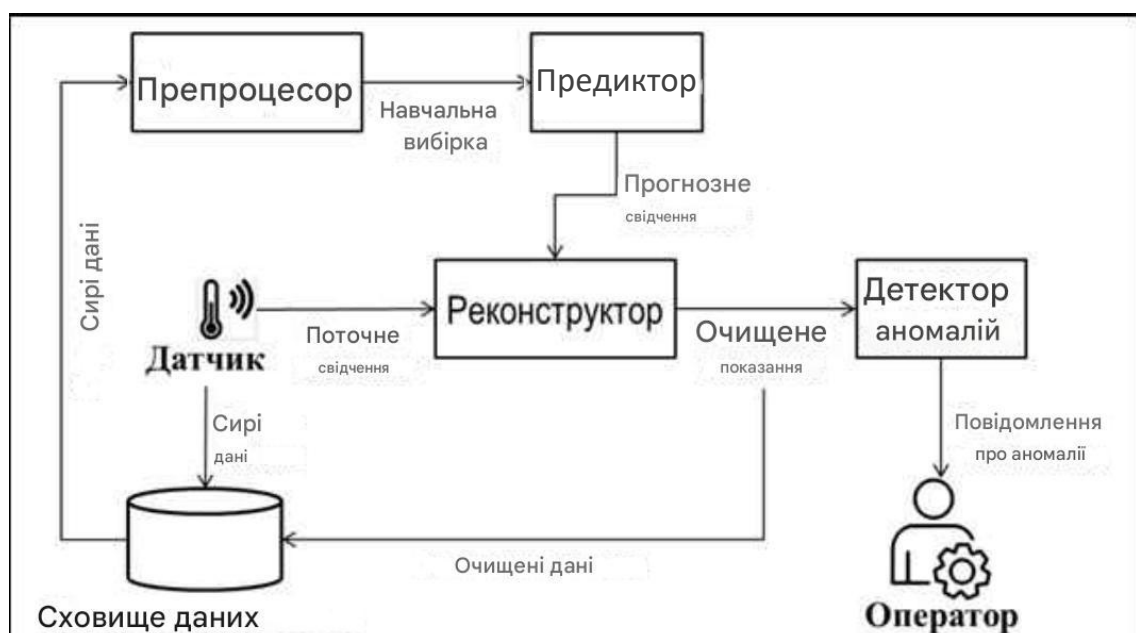


Рисунок 2.1 – Архітектура Модуля очищення даних [29]

Технологічний цикл Модуля очищення даних для деякого датчика виглядає так. Препроцесор виконує свої дії регулярно з частотою, що визначається оператором системи (типovим випадком є запуск препроцесора один раз на місяць). Препроцесор витягує зі сховища даних неочищені показання датчика, накопичені до поточного моменту, і готує навчальну вибірку нейронної мережі підсистеми Предиктор.

Предиктор запускається регулярно, але відповідно до частоти зняття показань датчика (наприклад, один раз на 15 хвилин), і виконує наступні дії. Нейронна мережа прогнозує значення датчика. Якщо при цьому датчик не зняв показання, воно замінюється на прогнозне. Інакше підсистема Реконструктор виконує бінарну класифікацію значення, одержаного від датчика: викид чи норма. Якщо поточне значення розпізнається як викид, Реконструктор замінює поточне значення спрогнозоване. Модуль очищення даних передає спрогнозовані значення системі збереження в сховищі даних.

Наприкінці циклу Детектор аномалій визначає, чи завершує поточне значення датчика деяку аномальну послідовність значень датчика, і повідомляє оператора системи у разі.

Далі Препроцесор розглянуто докладніше.

2.2. Опис Препроцесора

Препроцесор виконує підготовку навчальної вибірки для нейронної мережі підсистеми Предиктор і складається з наступних основних підсистем:

- парсер;
- відновник;
- детектор викидів;
- нормалізатор.

Технологічний цикл роботи Препроцесора представлений на рис. 2.2.

Спочатку Парсер отримує дані датчика зі сховища даних і перетворює їх у формат, придатний для подальшої обробки. Відновник замінює пропущені

значення на правдоподібні синтетичні. Після цього Детектор викидів знаходить у даних точки-викиди та замінює їх порожніми значеннями «NULL». Отримані дані повторно подаються на вхід Відновника, який замінює викиди на правдоподібні (синтетичні) значення. На останньому кроці Нормалізатор формує з отриманих даних набір нормалізованих підпоследовностей, які є навчальною вибіркою нейронної мережі підсистеми Предиктор.



Рисунок 2.2 – Архітектура Препроцесора

Відновник використовує сезонну авторегресійну інтегровану модель ковзного середнього з екзогенними регресорами SARIMAX (Seasonal AutoRegressive Integrated Moving Average with exogenous regressors) [28], яка часто застосовується для прогнозування значень сезонних часових рядів у різних предметних областях. Властивість сезонності часового ряду має на увазі наявність у цьому ряді коливань, що виникають із деякою періодичністю. Для застосування SARIMAX у цій предметній області необхідно підібрати такі параметри:

- p – порядок авторегресії;
- d – порядок інтегрування;
- q – порядок ковзного середнього;

- P – сезонний порядок авторегресії;
- D – порядок сезонного інтегрування;
- Q – сезонний порядок ковзного середнього;
- s – період сезонності часового ряду.

Застосування SARIMAX можливо, якщо цільовий часовий ряд є стаціонарним. Під стаціонарними часовими рядами розуміють такі часові ряди, елементи яких є випадковими величинами з постійним математичним очікуванням та постійною дисперсією. Стаціонарність часового ряду перевіряється за допомогою наступних двох тестів, що застосовуються послідовно: розширений тест Дікі-Фуллера (ADF) [29] та тест Квятковського-Філіпса-Шмідта-Шина (KPSS). Якщо зазначені тести не підтвердять стаціонарність часового ряду, то для застосування моделі SARIMAX необхідно перетворити елементи ряду залежно від його характеристик, наприклад, логарифмувати елементи у разі наявності в ряді мультиплікативної сезонності (істотної відмінності дисперсії елементів у різних ділянках ряду).

У формальному викладі робота Препроцесора виглядає так. Парсер витягує зі сховища даних тимчасовий ряд T , який є хронологічно впорядкованою послідовністю числових чи порожніх значень: $T_{i,n} = (t_1, \dots, t_m)$, де $t_i \in \mathbb{R}$ або $t_i = \text{Null}$. Число m – довжина часового ряду. Потім тандем Відновника та Детектора викидів перетворює T таким чином, щоб $\forall i T_i \in \mathbb{R}$.

Далі Нормалізатор готує навчальну вибірку, кожен елемент якої є парою «послідовність показань датчика» і «прогнозне значення датчика». Як перший елемент зазначеної пари береться нормалізована підпослідовність показань датчика фіксованої довжини, як другий елемент – одне нормалізоване показання датчика, наступне за даною підпослідовністю.

Під послідовністю $T_{i,n}$ часового ряду назвемо безперервний проміжок значень зазначеного ряду, що складається з n елементів і починаються з позиції i : $T_{i,n} = (t_1, \dots, t_{i+n-1})$, $1 \leq i \leq m-n+1$. Використовуючи мінімаксну нормалізацію, для підпослідовності $T_{i,n}$ її нормалізована версія $\tilde{T}_{i,n}$ обчислюється як:

$$\tilde{T}_{i,n} = (\tilde{t}_i, \dots, \tilde{t}_{i+n-1}), \quad (2.1)$$

$$\text{де } t_i = \frac{t_i - t_{\min}}{t_{\max} - t_{\min}}.$$

Для нормалізованою підпоследовності $\tilde{T}_{i,n}$ елемент \tilde{t}_i розглядається як її прогнозне значення. Для подальшого використання введемо позначення S_T^n набору нормалізованих підпоследовностей тимчасового ряду T довжину n ($n \ll m$), і за P позначимо набір відповідних цим підпоследовності прогнозів.

Довжина підпоследовності n ($n \ll m$) є параметром Модуля очистки і обчислюється як $n = \text{frequency} * \text{horizon}$, де frequency - частота датчика, а horizon – історичний горизонт (довжина тимчасового інтервалу в минулому), що використовується. Предиктором та підбирається оператором SCADA-системи «ПоліТЕР». Наприклад, типова частота зняття показань температурного датчика 4 рази на годину, історичний горизонт обраний рівним 12 годин, тоді довжина підпоследовності, використовувана Модулем очищення, становить $n=48$.

2.3 Функціональні та нефункціональні вимоги

Функціональні вимоги – це вимоги, які визначають дії, які має виконувати система, не враховуючи обмежень, що з її реалізацією, тобто визначають поведінка системи у процесі обробки інформації. З наведеного вище опису зазначимо основні вимоги до системи.

Стема повинна:

- 1) завантажити дані із зовнішнього джерела та приводити їх до зручного для обробки вигляду;
- 2) повертати дані навчальної вибірки;
- 3) відновлювати свідчення, що мають значення NULL;
- 4) знаходити дані, які вважаються короточасними викидами (визначимо їх як 1-3 значення, що мають значну відмінність від последовності)

і надавати їм значення NULL;

5) нормалізувати весь часовий ряд.

Нефункціональні вимоги – це вимоги, які визначають поведінка системи, але описують атрибути системи чи атрибути системного оточення.

При описі системи було виявлено основна дисфункція - система повинна мати помилку відновлення не менше 0,2, розрахована за RMSE (середньоквадратична помилка).

2.4 Варіанти використання системи

Для проектування програми було використано мову графічного опису об'єктного моделювання UML [30].

Було побудовано модель взаємодії актора «Зовнішня система» («external system») із Препроцесором. Діаграма варіантів використання (use-case diagram) представлена на рис. 2.3.

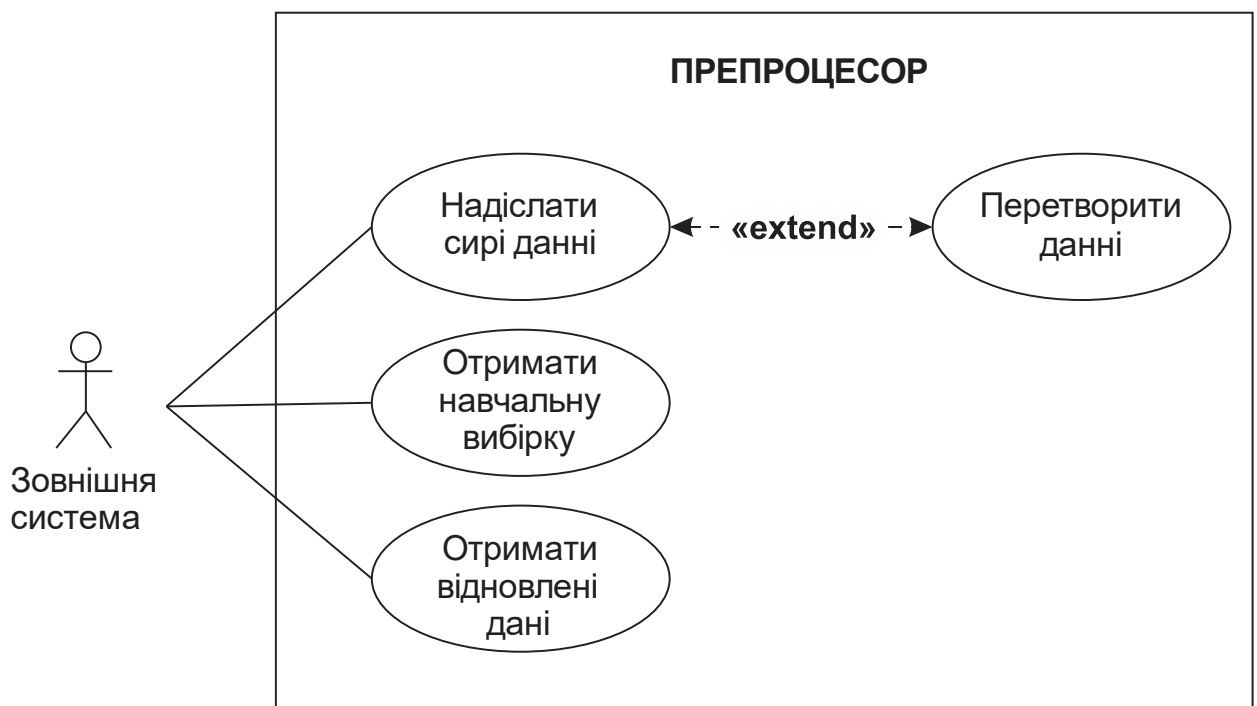


Рисунок 2.3 – Діаграма варіантів використання

Актором є зовнішня система (external system) на представленій діаграмі, яка має чотири основні варіанти використання системи:

- 1) надіслати сирі дані;
- 2) отримати навчальну вибірку;
- 3) отримати відновлені дані;
- 4) прецедент перетворити дані, що включається в перший варіант використання.

Специфікація варіантів використання, наведено у додатку А.

2.5 Діаграма компонентів для Препроцесора

У цьому розділі розглядається розроблена діаграма компонентів, яка показує розбиття системи на структурні компоненти та зв'язки між ними. Розроблена діаграма представлена на рис. 2.4.

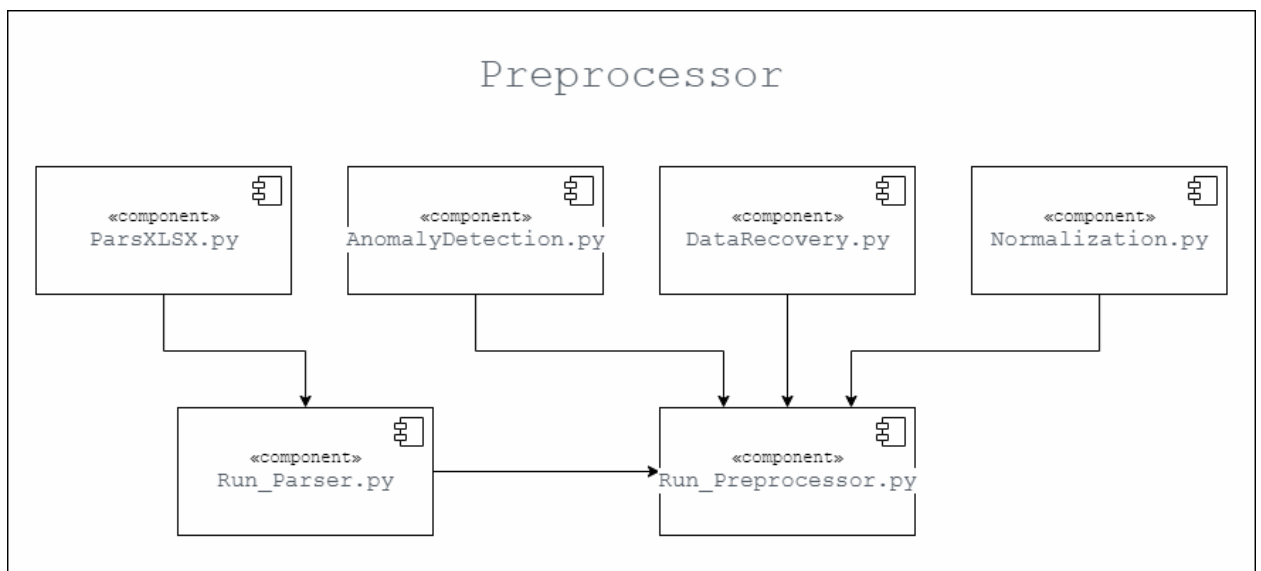


Рисунок 2.4 – Діаграма компонентів

Ця діаграма складається з наступних компонентів:

- 1) ParsXLSX.py – експортує функцію читання та парсингу даних із

файлу формату*.xlsx, з параметрами, що налаштовують парсер залежно від структури файлу;

2) AnomalyDetection.py- Експортує функцію для пошуку точкових аномалій (викидів);

3) DataRecovery.py- Експортує функцію для відновлення пропущених значень;

4) Normalization.py- Експортує функцію нормалізації даних;

5) Run_Parser.py – виконує функцію читання та парсингу даних із файлу формату*.xlsx, формує директорію та наповнює її залежно від наповнення файлу (аркуш файлу – каталог в директорії, стовпець – файл *.csv);

6) Run_Preprocessor.py – виконує функції читання заданого файлу, пресингу даних, пошуку точкових аномалій (викидів), відновлення пропущених значень та нормалізації часового ряду.

2.6 Діаграма діяльності для Препроцесора

Відповідно до вимог, які надані вище у розділі, реалізовано діаграму діяльності, яка показує, як відбувається процес взаємодії зовнішньою системою та препроцесорами. На рисунку 2.5 представлено діаграму діяльності.

У представленій діаграмі зовнішня система звертається до функцій препроцесора.

Система надсилає команду на зчитування файлу. Відповідно до встановлених конфігураційних параметрів, Препроцесор стартує парсинг, який перевіряє коректність даних, та їх цілісність. Якщо дані коректні і в них присутні перепустки та викиди, запускається перетворення даних. Результат перетворення повертається лише після завершення роботи Препроцесора з очищення від викидів та відновлення даних, які записуються у файл формату *.csv. Потім запускається нормалізація, яка формує, як і на попередньому кроці *.csv, але з позначкою normalized.

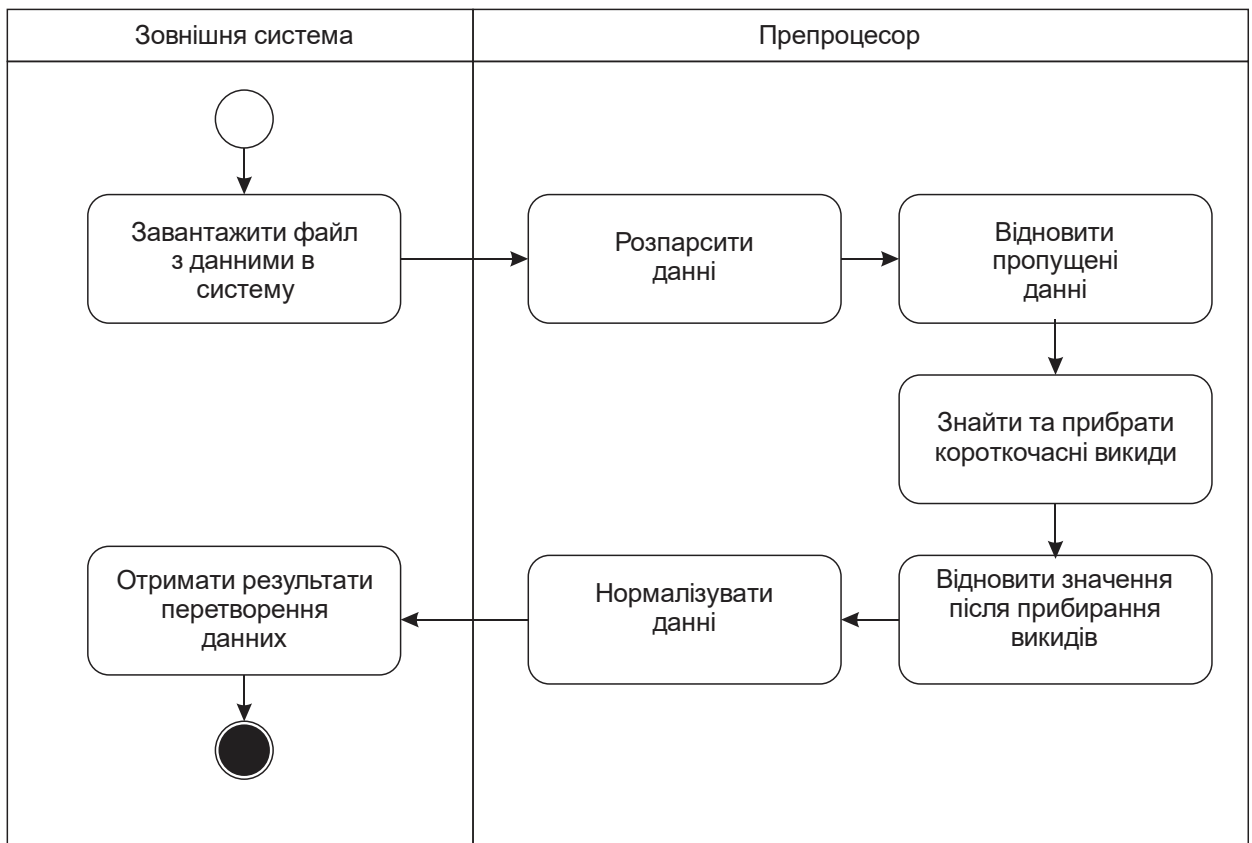


Рисунок 2.5 – Діаграма діяльності

Після всього перетворення зовнішня система може користуватися даними, отриманими в результаті роботи Препроцесора, що відповідає дії «Отримати результати перетворення даних».

Висновки до другого розділу

У другому розділі спроектовано архітектуру модуля очищення даних та детально описано структуру препроцесора. Визначено основні компоненти системи та їх взаємодію. Сформульовано функціональні та нефункціональні вимоги до системи, зокрема щодо точності обробки даних.

Розроблено UML-діаграми, які відображають логіку роботи системи та взаємодію з зовнішніми компонентами. Обґрунтовано використання моделі

SARIMAX для відновлення часових рядів і методів нормалізації даних для підготовки навчальних вибірок.

РОЗДІЛ 3 ПРОГРАМНЕ ТА ТЕХНІЧНЕ ЗАБЕЗПЕЧЕННЯ

3.1 Програмні засоби реалізації

Препроцесор реалізований за допомогою бібліотек мови Python версії 3.10 [31].

Парсер використовує стандартні бібліотеки `orenpurxl` версії 3.0.10 та `pandas` версії 1.3.6 [32].

Детектор викидів ґрунтується на алгоритмах бібліотеки `adtk` версії 0.6.2 (Anomaly Detection Toolkit).

Модель SARIMAX, що використовується Відновником, реалізована у стандартній бібліотеці моделей версії 0.13.2.

Нормалізатор реалізований за допомогою стандартної бібліотеки `sklearn` версії 1.1.1 [33].

Також використовувалися бібліотеки: `numpy` версії 1.18.4, `matplotlib` версії 3.1.1 і `unittest` версії 3.7 для проведення тестування [34].

3.2 Реалізація компонентів препроцесора

Містить три основні модулі:

- 1) `ParsXLSX`;
- 2) `AnomalyDetection`;
- 3) `DataRecovery`;
- 4) `Normalization`.

Для виконання наведених завдань необхідно встановити додаткові пакети Python.

Встановлення Пакета `statsmodels` виконується через командний рядок (`cmd`) системи: представлено на рис. 31.

```
git clone git: //github.com/statsmodels/statsmodels.git cd statsmodels
python setup.py build
python setup.py install
```

Рисунок 3.1 – Підготовка до та встановлення пакету statsmodels

Інші бібліотеки встановлюються аналогічно через командний рядок.

Файл config.py містить основні параметри, необхідні для роботи системи, а також їх детальний опис. Вигляд цього файлу наведено на рис. 3.2 - 3.7.

Для опису значень кожної зміни були використані коментарі.

```
# ім'я файлу або шлях до файлу XLSX - excel file_name_xlsx = 'Дані
ВІМ_ВЕМ ЗБВ.xlsx'
# обов'язково вказувати розширення
# Ім'я папки для збереження даних у форматі CSV та
# Назва файлу, що містить імена сторінок
save_directory_CSV = 'SensorData' root_file_name = 'page_names.csv'
# обов'язково вказувати розширення
# файл з шляхами до даних з часовими рядами path_time_siries =
'pathFileSensors.csv'
```

Рисунок 3.2 – Конфігурація для опису шляхів у препроцесорі

Із лістингу видно, що для роботи системи необхідно вказати шлях до файлу, з якого будуть зчитуватися та перетворюватися дані.

```
# Вибір методу дедукції аномалій
# 0 - SearchOutliersADTK() - метод пошуку аномалій ADTK
# 1 - SearchOutliersMedian() - Пошук викидів anomaly_detect = 1
```

Рисунок 3.3 - Вибір алгоритму пошуку аномалій

Для роботи було реалізовано два алгоритми `searchOutliersADTK()` – метод пошуку аномалій з фільтрацією по групах, який дозволяє за допомогою угруповання оцінити, чи є ця точка викидом. `searchOutliersMedian()` – пошук викидів за медіанним значенням, коли аномальною точкою вважається та, чие відхилення свідомо більше медіанного значення.

```

searchOutliersADTK()
# Детекція аномалій
#
# SearchOutliersADTK(time_siries, freq=365, c=7.0, n=3)
#   freq - сезонна варіативність число від 12 до
#   len(time_siries)/2. Default: 365
#   side - Зміна перепаду
#   “both” – в обидві сторони;
#   “positive” - зростання
#   “negative” - спадання
#   c - Коефіцієнт змінності викиду.   Default: 7.0
#   n - фільтр точкової аномалії
# (викидом вважати то кількість точок поспіль яка менше n)
SearchOutliersADTK_freq = 365
SearchOutliersADTK_c = 7.0
SearchOutliersADTK_n = 3

```

Рисунок 3.4 – Характеристика алгоритмів пошуку викидів для алгоритму

```

# Відновлення пропущених значень
#
#   order:
#   d - має бути цілим числом, що вказує порядок
# інтегрування процесу Default: 1,
#   p, q - можуть бути або цілими числами, що вказують
# порядки AR та MA,
# або ітераціями дають певний AR та/або MA
# Default: p = 2, q = 4
order_d = 1
order_p = 2
order_q = 4
# seasonal_order:
#
#   D - має бути цілим числом, що вказує порядок
# інтегрування процесу Default: 1
#   P, Q - можуть бути або цілими числами, що вказують
# порядки AR і MA чи ітераціями
#   дають конкретні AR та / або MA
#   Default: p = 2, q = 3
#   s - ціле число, що задає періодичність
# (кількість періодів у сезоні)
часто воно дорівнює 4 для квартальних даних
#   або 12 для місячних даних тощо
seasonal_order_D = 1
seasonal_order_P = 2
seasonal_order_Q = 3
seasonal_order_s = 12

```

Рисунок 3.5 – Характеристика алгоритму відновлення пропущених значень

```

searchOutliersMedian()
# SearchOutliersMedian(time_series, c=3, dev=0.02):
#     c - Коефіцієнт змінності викиду.   Default: 3.0
#     dev - поправочний коефіцієнт варіативності
# (залежить від середнього значення датчика)
# Вихідні дані:
#     outlier - pd.DataFrame(),
# формату: Datetime;
# value(True - аномалія False - відсутність аномалії)
SearchOutliersMedian_c = 3.0
SearchOutliersMedian_dev = 0.02

```

Рисунок 3.6 – Характеристика алгоритмів пошуку викидів для алгоритму

```

# AnomalyDetection name_file_AnomalyDetectionData =
r"Tests\Data\Anomaly\Data\ec2_cpu_utilization_5f5533.csv"
name_file_AnomalyDetectionTest =
r"Tests\Data\Anomaly\Labels\ec2_cpu_utilization_5f5533Test.csv"
# DataRecovery
name_file_DataRecoveryData = r>Data\Predict\Train1.csv"
name_file_DataRecoveryTest = r>Data\Predict\Test1.csv"

```

Рисунок 3.7 – Шляхи для тестових даних

Даний конфігураційний файл дозволяє налагоджувати роботу препроцесора, проводити експерименти та аналіз для пошуку найбільш відповідних значень алгоритмів.

Для запуску препроцесора був розроблений спеціальний пусковий файл, який містить виклики реалізованих компонентів системи в хронологічному порядку.

3.3 Опис модульного тестування препроцесора

Тести, що використовуються для перевірки системи, розподілені по чотирьох файлах, кожен з яких виконує тестування одного компонента з архітектури препроцесора мають такі назви файлів:

- 1) AnomalyDetectionTest.py;
- 2) DataRecoveryTest.py;
- 3) ParsXLSXTest.py;
- 4) NormalizationTest.py.

Кожен із цих файлів тестує всі експортовані функції з однойменних модулів препроцесора.

AnomalyDetectionTest.py тестує експортовані функції AnomalyDetection.py.

Дані цієї функції представлені на рис. 3.8, а вихідні очікувані дані в рис. 3.9, де:

- 1.0 – аномалія;
- 0.0 – відсутність аномалії.

Datetime, value

2026-04-07 01:50:00,19.6866160273

2026-04-07 01:55:00,19.3563053672

2026-04-07 02:00:00,20.4140421657

2026-04-07 02:05:00,18.8776063185

Рисунок 3.8 – Вхідні дані для тестування функції AnomalyDetection()

Datetime, value
2026-04-07 01:50:00; 0.0
2026-04-07 01:55:00; 0.0
2026-04-07 02:00:00; 1.0
2026-04-07 02:05:00; 0.0

Рисунок 3.9 – Вихідні дані тестування функції AnomalyDetection()

Тестування DataRecovery, Normalization і ParsXLSX відбувається за рахунок перевірки на невеликому наборі даних, із заздалегідь розрахованим результатом, який перевіряє основні позитивні та негативні кейси.

3.4 Обчислювальні експерименти

3.4.1 Набор даних

Для оцінки запропонованого підходу було проведено обчислювальні експерименти на реальних даних, взятих зі сховища даних SCADA-системи «ПоліТЕР». Як дані використовувалися показання температурного датчика. Частота зняття показань датчика становить 4 рази на годину.

Після обробки зазначених даних Препроцесором перші 42 тижні року (80%) показань були використані в двох якостях: як вибірка для підбору параметрів моделі SARIMAX Відновника. Інші 8 тижнів року (20 %) показань датчика, що залишилися, моделювали роботу модуля очищення даних у штатному режимі і використовувалися як тестова вибірка.

3.4.2 Проведення експериментів для алгоритму відновлення

Для перевірки точності та коректності роботи алгоритму відновлення

було підібрано параметри для моделі SARIMAX [43]:

- порядок авторегресії $p = 0, 7$;
- порядок інтегрування $d = 1$;
- порядок ковзного середнього $q = 0, 7$;
- сезонний порядок авторегресії $P = 1$;
- порядок сезонного інтегрування $D = 1$;
- сезонний порядок ковзного середнього $Q = 0$;
- період сезонності часового ряду $S = 96$.

Що дозволило отримати максимальну точність відновлення даних у Препроцесорі.

На рис. 3.10 представлено графік прогнозу на один день, отриманий за допомогою моделі SARIMAX (6, 1, 5) (0, 1, 1) 96 на різних датчиках, але в один проміжок часу.

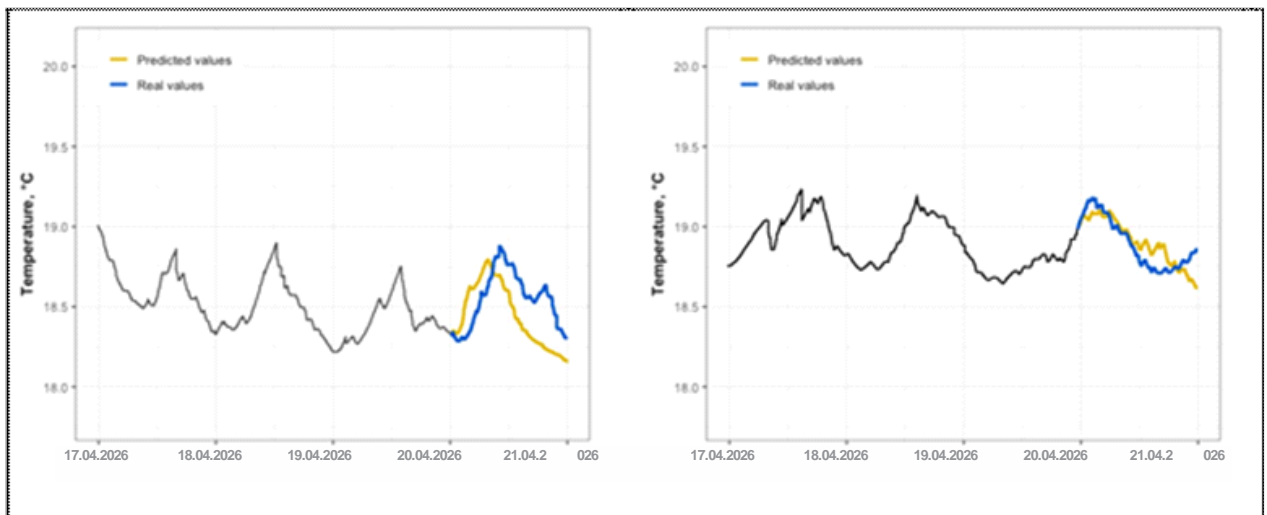


Рисунок 3.10 – Прогноз на один день, отриманий за допомогою моделі SARIMAX (6, 1, 5) (0, 1, 1) 96 для двох різних місць

По представленим графікам вище була розрахована середньоквадратична помилка:

- а) RMSE 0,183;
- б) RMSE 0,112.

На рис. 3.11 прогнозу на один день, отриманої на одному датчику в різний період часу.

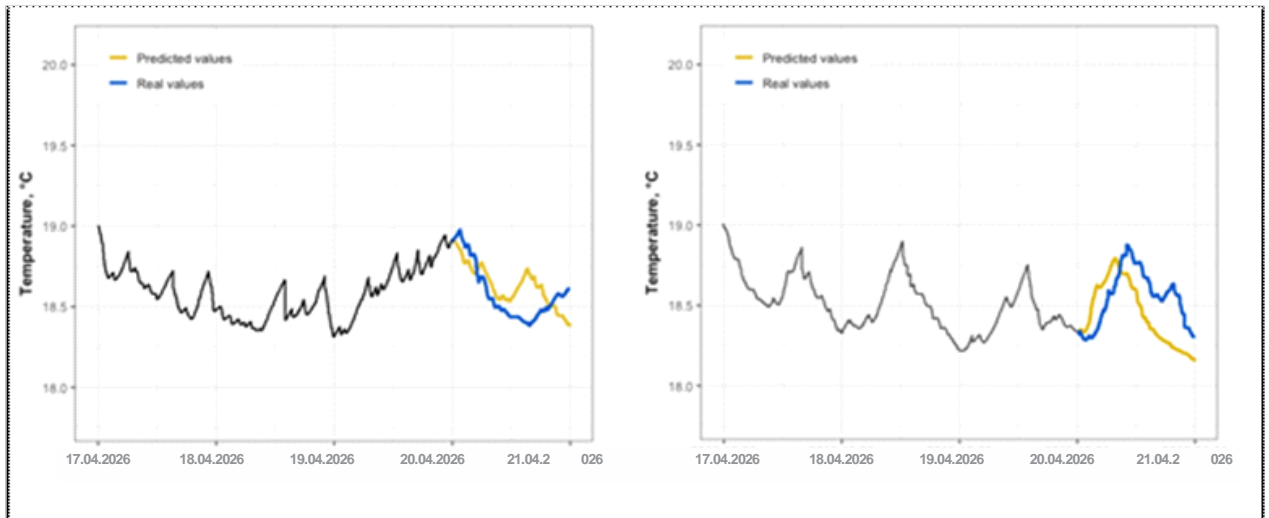


Рисунок 3.11 – Прогноз на один день, отриманий за допомогою моделі SARIMAX (6, 1, 5) (0, 1, 1) 96 для двох проміжників часу

По представленим графікам вище була розрахована середньоквадратична помилка:

- а) RMSE 0,116,
- б) RMSE 0,121.

На рис. 3.12 представлено прогноз на один день, а також побудовано 80% та 95% довірчі інтервали прогнозу.

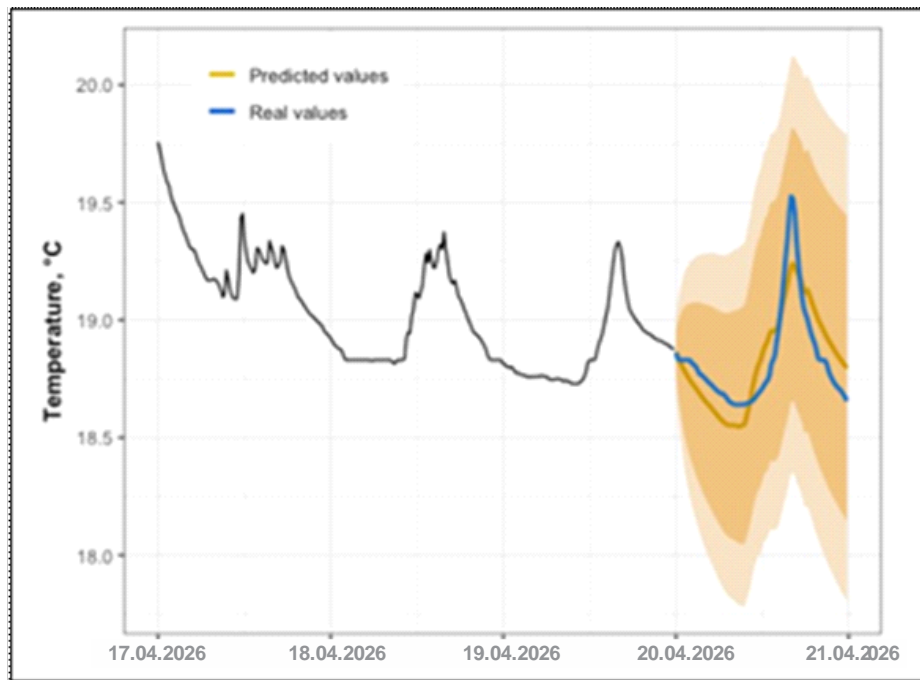


Рисунок 3.12 – Прогноз на один день, отриманий за допомогою моделі SARIMAX (6, 1, 5) (0, 1, 1) 96

Середня помилка прогнозу на наборі даних при проведенні численних експериментів (більше 10 датчиків температури, при відновленні трьох різних ділянок для однієї) становить принаймні RMSE 0,129, що відповідає вимогам, зазначеним у пункті 3.3 цієї роботи (RMSE трохи більше 0,2).

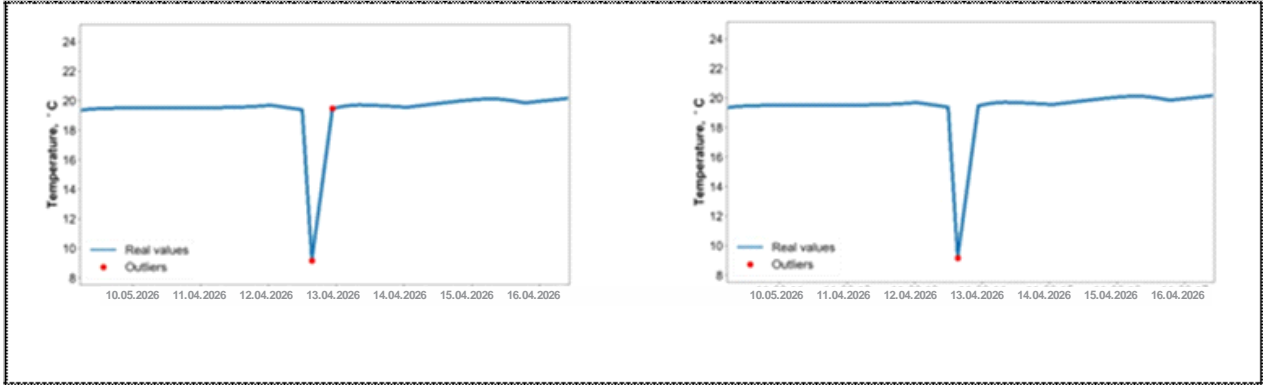
3.5 Проведення експериментів для алгоритму пошуку викидів

Для перевірки роботи алгоритму пошуку викидів були проведені експерименти, з яких було зроблено висновок, що алгоритм відповідає вимогам, поставленим до Препроцесора. При розробці до експериментів було запропоновано два алгоритми:

- SearchOutliersADTK() –метод пошуку аномалій ADTK із бібліотеки Anomaly Detection Toolkit (adtk) [35];
- SearchOutliersMedian() –Пошук викидів шляхом розрахунку різниці з медіаною ряду.

Демонстрація роботи цих алгоритмів представлена на рис. 3.13 (а – ADTK, б – Median) у якому червоними точками позначено точка, яку алгоритм

ВВАЖАВ ВИКИДОМ.

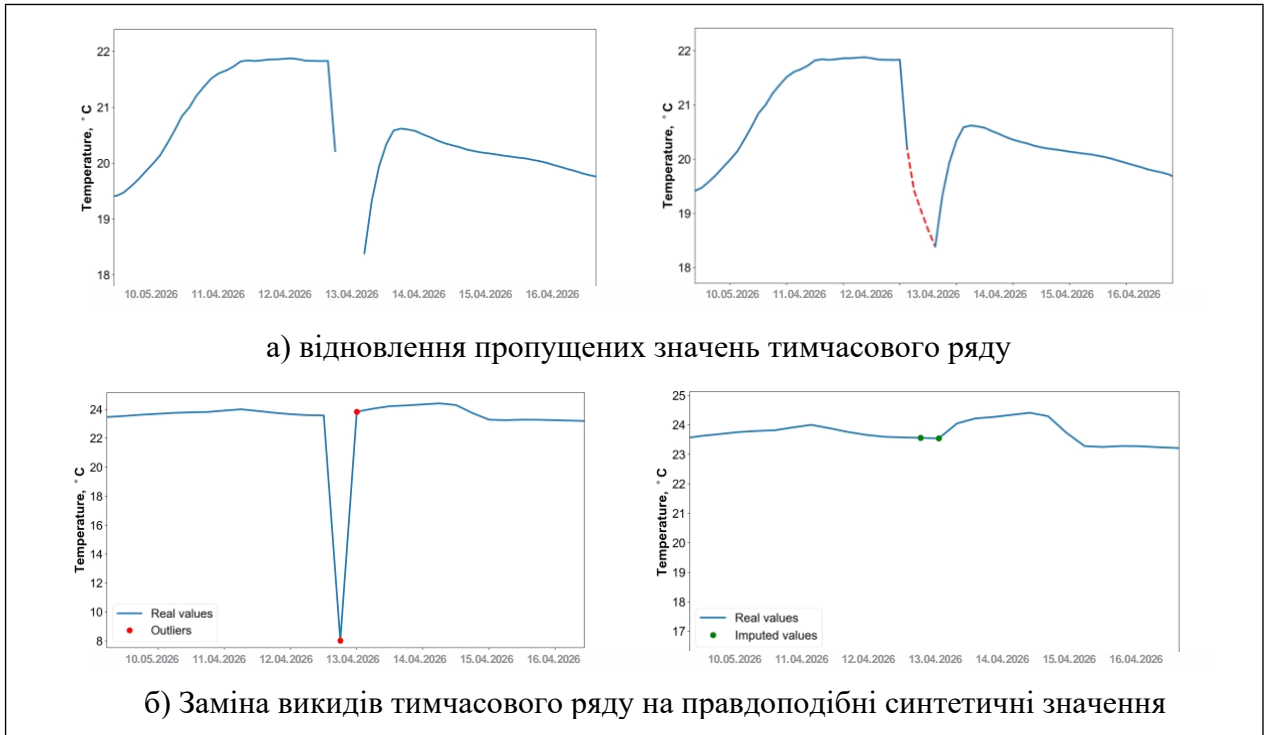


а) ADTK

б) Median

Рисунок 3.13 – Заміна викидів тимчасового ряду на правдоподібні синтетичні значення

Результати роботи Препроцесора з відновлення та видалення аномалій наведено на рис. 3.14.



а) відновлення пропущених значень тимчасового ряду

б) Заміна викидів тимчасового ряду на правдоподібні синтетичні значення

Рисунок 3.14 – Результати роботи Відновлювача на основі моделі SARIMAX (1) (0 1 1) та алгоритмом відновлення ADTK

Наведені алгоритми відповідають функціональним вимогам. Для постійної роботи Препроцесора було обрано алгоритм ADTK. Так як на більшому наборі даних під час тестування він показав більшу точність роботи.

Висновок до третього розділу

У третьому розділі реалізовано програмну систему препроцесінгу даних із використанням сучасних бібліотек Python. Розроблено основні модулі системи та проведено їх модульне тестування. Виконано обчислювальні експерименти на реальних даних, які підтвердили ефективність запропонованих алгоритмів.

Досягнуто необхідної точності відновлення даних ($RMSE < 0,2$), що відповідає встановленим вимогам. Обґрунтовано вибір алгоритму виявлення аномалій (ADTK) як найбільш ефективного. Отримані результати підтверджують працездатність і практичну цінність розробленої системи.

РОЗДІЛ 4 ОХОРОНА ПРАЦІ

4.1 Організаційно-правові основи забезпечення безпеки праці

Охорона праці є важливою складовою організації трудової діяльності в сучасних умовах. Вона спрямована на створення безпечних і комфортних умов праці, збереження життя, здоров'я та працездатності працівників у процесі виконання професійних обов'язків. Особливого значення питання охорони праці набувають у сфері інформаційних технологій та автоматизованих систем керування, де працівники тривалий час працюють із комп'ютерною технікою, програмним забезпеченням та електронним обладнанням.

У даній роботі об'єктом проектування виступає автоматизоване робоче місце оператора системи керування опаленням будівлі. Робота оператора пов'язана з використанням персонального комп'ютера, програмних комплексів для аналізу та обробки даних, а також засобів контролю параметрів системи опалення. Під час виконання професійних обов'язків працівник може піддаватися впливу різних небезпечних і шкідливих факторів, серед яких електричний струм, недостатнє освітлення, нервово-емоційне навантаження, тривала статична поза та напруження зору. Саме тому забезпечення належних умов праці є необхідною умовою ефективної та безпечної роботи оператора.

Організація охорони праці в Україні регулюється законодавчими та нормативно-правовими актами. Основним документом у цій сфері є Закон України «Про охорону праці», який визначає основні права та обов'язки працівників і роботодавців щодо створення безпечних умов праці [36]. Також питання охорони праці регулюються Кодексом законів про працю України, державними санітарними нормами, правилами електробезпеки та іншими нормативними документами, що відповідають специфіці виконуваних робіт.

Державна політика у сфері охорони праці спрямована на запобігання виробничому травматизму та професійним захворюванням, забезпечення належного рівня безпеки працівників, проведення профілактичних заходів та

контроль за дотриманням вимог безпеки [36]. Важливу роль також відіграють міжнародні стандарти у сфері безпеки праці, зокрема стандарт ДСТУ ISO 45001:2019, який встановлює вимоги до систем управління охороною здоров'я та безпекою праці.

Для безпечної експлуатації автоматизованого робочого місця оператора системи керування опаленням необхідно забезпечити правильну організацію робочого простору, відповідний мікроклімат приміщення, достатній рівень освітлення, дотримання правил електробезпеки та пожежної безпеки та інших параметрів. Також важливим є дотримання режиму праці та відпочинку під час роботи за комп'ютером, що дозволяє знизити рівень втоми та запобігти негативному впливу на здоров'я працівника.

4.2 Характеристика об'єкта та виявлення потенційних небезпек

Об'єктом проектування у розділі з охорони праці виступає автоматизоване робоче місце оператора системи керування опаленням будівлі. Дане робоче місце призначене для контролю, обробки та аналізу даних, які надходять від системи опалення та датчиків температури. За допомогою програмного забезпечення оператор здійснює моніторинг параметрів роботи системи, контролює температурний режим у приміщеннях, аналізує отримані дані та приймає рішення щодо коригування режимів роботи обладнання.

Автоматизоване робоче місце включає персональний комп'ютер, монітор, мережеве обладнання, програмне забезпечення для обробки даних та засоби підключення до системи керування опаленням. Робота системи заснована на постійному обміні інформацією між датчиками, сервером та оператором. Основною особливістю функціонування даного об'єкта є безперервна обробка інформації та необхідність постійного контролю за станом системи опалення.

Експлуатація автоматизованого робочого місця здійснюється в офісному приміщенні з використанням електронного обладнання та комп'ютерної

техніки. Працівник проводить значну частину робочого часу за персональним комп'ютером, що супроводжується тривалим навантаженням на зір, нервову систему та опорно-руховий апарат. До категорії персоналу, який працює з даним об'єктом, належать оператори систем керування, інженери та технічні спеціалісти.

Під час роботи на автоматизованому робочому місці можуть виникати різні небезпечні та шкідливі виробничі фактори [37, 38]. Основними джерелами потенційної небезпеки є комп'ютерне обладнання, електрична мережа, тривала робота за монітором та несприятливі умови праці у приміщенні.

Таблиця 4.1 – Результати виявлення небезпек стосовно об'єкту проектування

| № | Потенційна небезпека | Джерело небезпеки | Можливі наслідки |
|---|-------------------------------|---|--|
| 1 | Ураження електричним струмом | Комп'ютерне та мережеве обладнання | Травми, ураження струмом |
| 2 | Пожежна небезпека | Перевантаження електромережі, несправність обладнання | Загоряння обладнання, пошкодження техніки |
| 3 | Напруження зору | Тривала робота за монітором | Втома очей, погіршення зору |
| 4 | Статичне фізичне навантаження | Тривале перебування у сидячому положенні | Біль у спині, порушення постави |
| 5 | Нервово-емоційне навантаження | Постійний контроль та аналіз інформації | Втома, зниження працездатності |
| 6 | Недостатнє освітлення | Невідповідні умови освітлення робочого місця | Швидка втомлюваність, перенапруження зору |
| 7 | Перегрів обладнання | Безперервна робота технічних засобів | Збої у роботі системи, вихід обладнання з ладу |

Таким чином, під час експлуатації автоматизованого робочого місця оператора системи керування опаленням існує ряд небезпечних та шкідливих факторів, які можуть негативно впливати на стан здоров'я працівника та надійність роботи обладнання. Для зниження ризику виникнення небезпечних ситуацій необхідно передбачити відповідні організаційні та технічні заходи безпеки.

4.3 Дослідження ризику реалізації потенційних небезпек на об'єкті проектування та розробка заходів щодо їх попередження

Дослідження ризику реалізації потенційних небезпек є важливою складовою системи охорони праці на будь-якому підприємстві або робочому об'єкті. Основною метою такого дослідження є своєчасне виявлення небезпечних і шкідливих факторів, оцінювання рівня їх впливу на працівників та визначення ефективних заходів для запобігання виникненню небезпечних ситуацій [39]. Проведення оцінювання ризиків дозволяє не лише підвищити рівень безпеки праці, а й знизити ймовірність виникнення аварій, професійних захворювань та травматизму.

Для автоматизованого робочого місця оператора системи керування опаленням питання дослідження ризиків є особливо актуальним, оскільки робота пов'язана з використанням електронного обладнання, персонального комп'ютера та програмних комплексів обробки даних. Під час експлуатації такого обладнання працівник може піддаватися впливу різних небезпечних факторів, які здатні негативно впливати як на стан здоров'я людини, так і на працездатність технічних засобів.

Важливість оцінювання ризиків полягає у тому, що навіть незначні порушення умов праці або несправності обладнання можуть призвести до серйозних наслідків [39]. Наприклад, несправність електромережі або перевантаження обладнання можуть стати причиною короткого замикання чи пожежі. Недостатній рівень освітлення та тривала робота за монітором можуть

викликати перевтому, погіршення зору та зниження працездатності оператора. Тривале перебування у сидячому положенні створює додаткове навантаження на опорно-руховий апарат, що може призвести до розвитку професійних захворювань.

Дослідження ризику реалізації потенційних небезпек дозволяє визначити найбільш небезпечні фактори на робочому місці, оцінити ймовірність їх виникнення та можливі наслідки для працівника і обладнання [39]. На основі отриманих результатів розробляються організаційні та технічні заходи, спрямовані на зниження рівня ризику та покращення умов праці.

Тобто, дослідження ризиків є необхідною умовою забезпечення безпечної експлуатації автоматизованого робочого місця оператора системи керування опаленням. Проведення оцінювання ризиків та впровадження профілактичних заходів сприяє підвищенню рівня безпеки праці, збереженню здоров'я працівників та забезпеченню стабільної роботи системи керування опаленням.

Дослідження ризиків у розділі буде виконано методом аналізу дерева відмов, що дозволяє проаналізувати передумови реалізації досліджуваної небезпеки та своєчасно вжити заходів, спрямованих на її недопущення.

Результати побудови дерева для небезпек «ураження електричним струмом» та «статичне фізичне навантаження» представлені на рис. 4.1, 4.2.

Для зниження ризику ураження електричним струмом на автоматизованому робочому місці оператора системи керування опаленням необхідно впроваджувати комплекс організаційних та технічних заходів безпеки. Основна увага повинна приділятися забезпеченню справного технічного стану обладнання, дотриманню правил експлуатації електроприладів та підвищенню рівня підготовки персоналу.

Одним із найважливіших заходів попередження виникнення досліджуваної небезпеки є регулярна перевірка технічного стану електрообладнання [5]. Необхідно періодично контролювати стан кабелів, розеток, мережевих фільтрів та блоків живлення. У разі виявлення

пошкоджень ізоляції або ознак перегріву обладнання слід негайно припинити його експлуатацію та провести ремонт або заміну несправних елементів.

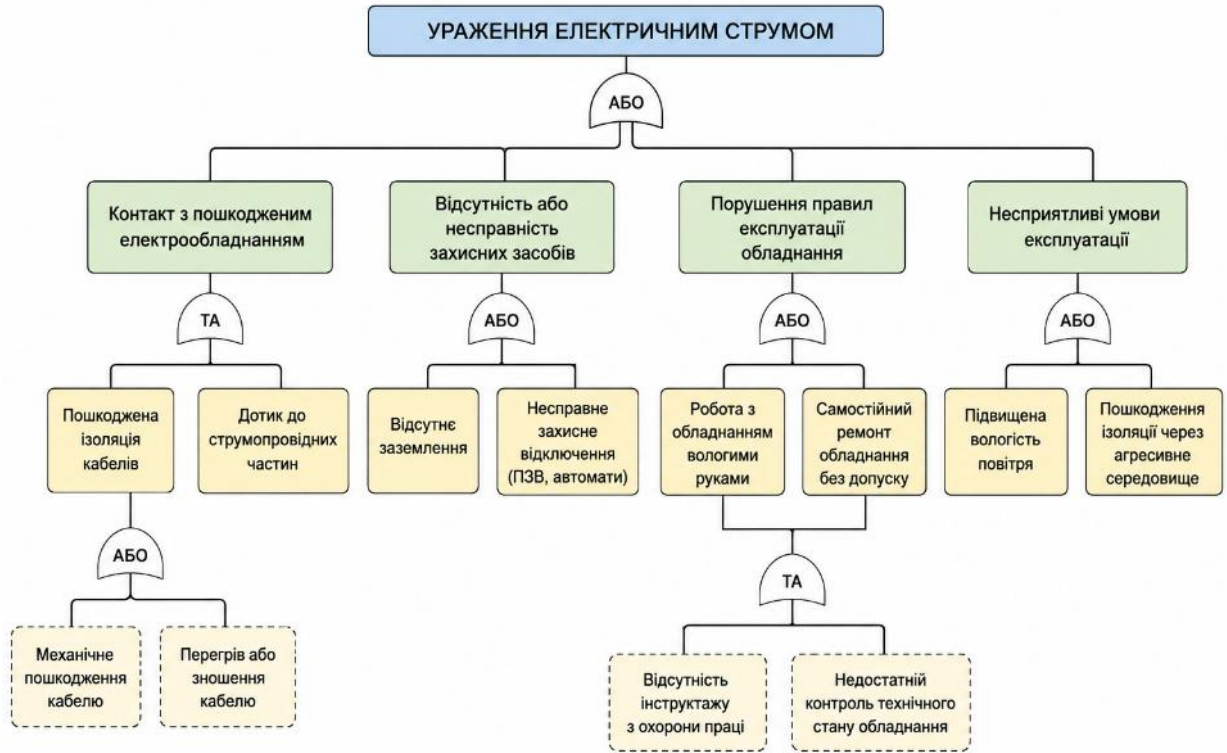


Рисунок 4.1 – Побудова дерева для небезпеки «ураження електричним струмом»

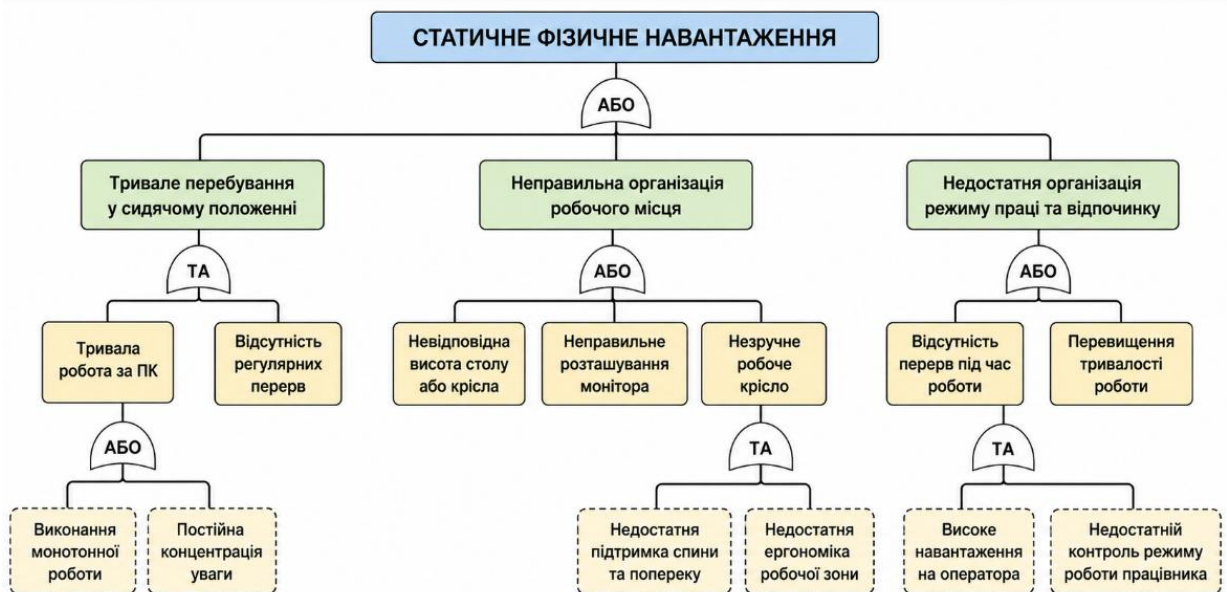


Рисунок 4.2 – Побудова дерева для небезпеки «статичне фізичне навантаження»

Для захисту працівників від ураження електричним струмом необхідно використовувати справну систему заземлення та автоматичні засоби захисного відключення. Використання автоматичних вимикачів і пристроїв захисного відключення дозволяє своєчасно припинити подачу електроенергії у випадку короткого замикання або витоку струму [40].

Важливе значення має правильна організація робочого місця. Електрообладнання повинно розташовуватися у сухому приміщенні з нормальним рівнем вологості та достатньою вентиляцією. Не допускається розміщення кабелів у місцях можливого механічного пошкодження або контакту з нагрівальними поверхнями.

Працівники повинні дотримуватися правил електробезпеки під час роботи з обладнанням. Забороняється працювати з технікою вологими руками, самостійно ремонтувати обладнання без відповідної кваліфікації або використовувати несправні електроприлади. Усі роботи з технічного обслуговування повинні виконуватися лише спеціалістами, які мають відповідний допуск [40].

Для мінімізації ризику виникнення аварійних ситуацій необхідно проводити регулярні інструктажі з охорони праці та електробезпеки. Працівники повинні бути ознайомлені з правилами користування обладнанням, порядком дій у разі виникнення несправностей та основними заходами надання першої допомоги при ураженні електричним струмом.

У разі виникнення небезпечної ситуації важливим є швидке реагування персоналу. На робочому місці повинні бути доступні засоби пожежогасіння, інструкції з евакуації та контакти аварійних служб. Це дозволить мінімізувати наслідки можливих аварій та забезпечити безпеку працівників.

Для зниження ризику виникнення статичного фізичного навантаження на автоматизованому робочому місці оператора системи керування опаленням необхідно забезпечити правильну організацію робочого процесу та створити комфортні умови праці. Основною причиною виникнення даної небезпеки є

тривале перебування працівника у сидячому положенні та постійна робота за персональним комп'ютером без достатнього відпочинку.

Одним із головних заходів профілактики є правильна організація робочого місця відповідно до ергономічних вимог. Робочий стіл та крісло повинні відповідати зросту працівника та забезпечувати зручне положення тіла під час роботи. Крісло має бути регульованим по висоті та мати підтримку для спини, що дозволяє зменшити навантаження на хребет і м'язи спини [40].

Важливу роль відіграє правильне розташування монітора та периферійного обладнання. Екран монітора повинен знаходитися на рівні очей працівника, а клавіатура та миша – у зручному положенні для рук. Це дозволяє знизити напруження м'язів шиї, плечей та рук під час роботи.

Для зменшення негативного впливу статичного навантаження необхідно дотримуватися режиму праці та відпочинку. Під час тривалої роботи за комп'ютером рекомендується робити короткі перерви через кожні 45-60 хвилин роботи. Під час перерв доцільно виконувати нескладні фізичні вправи, змінювати положення тіла та проводити розминку для м'язів спини, шиї та рук [41].

З метою профілактики перевтоми працівникам рекомендується чергувати роботу за комп'ютером з іншими видами діяльності, які не потребують постійного перебування у сидячому положенні [41]. Це сприяє зменшенню монотонності праці та покращує загальний фізичний стан працівника.

Необхідно також забезпечити належні умови мікроклімату у приміщенні. Комфортна температура повітря, достатній рівень освітлення та вентиляції позитивно впливають на самопочуття працівника та знижують рівень втоми під час роботи.

Важливим профілактичним заходом є проведення інструктажів та інформування працівників щодо правил організації робочого місця та дотримання ергономічних вимог. Працівники повинні знати про можливі наслідки тривалого статичного навантаження та способи його зменшення.

Висновок до четвертого розділу

У розділі «Охорона праці» було розглянуто особливості забезпечення безпечних умов праці для автоматизованого робочого місця оператора системи керування опаленням будівлі. Проведено аналіз організаційно-правових основ охорони праці, основних нормативних вимог та факторів, які впливають на безпеку працівника під час роботи з комп'ютерною технікою та електронним обладнанням.

У розділі надано характеристику об'єкта проектування та визначено основні потенційні небезпеки, що можуть виникати під час експлуатації автоматизованого робочого місця. Серед основних небезпечних та шкідливих факторів було виділено ураження електричним струмом, статичне фізичне навантаження, напруження зору, нервово-емоційне перевантаження та пожежну небезпеку.

Для найбільш значущих небезпек було проведено аналіз причин їх виникнення із використанням методу «дерева відмов». У результаті аналізу були визначені основні фактори, які можуть призводити до виникнення небезпечних ситуацій, а також встановлено взаємозв'язки між технічними та організаційними причинами ризиків.

Також у роботі були розроблені рекомендації щодо зниження рівня професійних ризиків та покращення умов праці. Запропоновані заходи спрямовані на підвищення рівня електробезпеки, покращення ергономічних умов праці, дотримання режиму праці та відпочинку, а також забезпечення належного технічного стану обладнання.

ЗАГАЛЬНІ ВИСНОВКИ

В рамках даної роботи було розроблено Препроцесор, який виконує функції обробки файлів формату *.xlsx, аналізу часових рядів на наявність викидів та перепусток, та застосування алгоритмів до них, для підготовки даних у подальшій роботі модуля очищення даних. Реалізовано алгоритми пошуку викидів та очищення даних від них та відновлення пропущених значень у часових рядах.

1. проведено огляд літератури та існуючих рішень з предметної галузі;
2. розроблено архітектуру модуля очищення даних та препроцесора;
3. виконано проектування системи;
4. розроблено та покрито модальними та інтеграційними тестами спроектована система;
5. проведено підбір параметрів роботи препроцесора та проведено експерименти.

У роботі порушено проблему очищення даних, що надходять з температурних датчиків в інтелектуальних системах опалення розумних будинків. Розглянуто випадок SCADA-системи "ПоліТЕР", встановленої в Південно-Уральському державному університеті для управління опаленням університетського кампусу. Представлено архітектуру та принципи реалізації модуля очищення даних, розробленого та впровадженого в систему «ПоліТЕР».

Для системи було проведено обчислювальні експерименти на реальних даних датчиків системи «ПоліТЕР», що показали високу точність роботи Модулю очищення даних та Препроцесора.

Таким чином, поставлені питання дослідження вирішені в повному обсязі, мета роботи - досягнута.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Дергачова В. В., Довгоруку О. М. Використання штучного інтелекту в енергоменеджменті промислових підприємств // Економічний вісник НТУУ «КПІ». – 2025. – № 33. – С. DOI: 10.20535/2307-5651.33.2025.335888
2. Яцків О. Ю., Коман Б. П. Визначення можливостей використання штучного інтелекту та машинного навчання для управління тепловими процесами в електронних пристроях // Technology Audit and Production Reserves. – 2025. – № 1. – С. —. – DOI: 10.15587/2706-5448.2025.323117
3. Ревуцька Н. Цифрові технології та енергоспоживання: вплив штучного інтелекту на кліматичні виклики // Інноваційне підприємництво: стан та перспективи розвитку : матеріали Міжнар. наук.-практ. конф. – Київ : КНЕУ, 2025. – С. 75–82
4. Сучасний розумний дім - Система розумного дому Електронний ресурс. URL: https://engocontrols.com/ua/?gad_source=1&gad_campaignid=23491594001&gbraid=0AAAABCqwfY-3bivdx52R7LnyV_wSAim&gclid=CjwKCAjw46HPBhAMEiwASZpLRClvmmo1KrkAt18AgusAHPapM7EzseffzafUVZjFPuzUuBgSvb3jRoC6kEQAvD_BwE (дата звернення 22.05.2026 р.).
5. Інтелектуальне опалення: Як сучасні технології полегшують контроль за температурою будинку Електронний ресурс. URL: <https://aqua-life.ua/ua/info/yntellektualnoe-otoplenye-kak-sovremennye-tekhnohyy-uproshchaiut-kontrol-za-temperaturoi-doma/?srsltid=AfmBOor7SG8dev1Gi7LcTvACA-5bACp6eJdDFcOZgKFnUnH1z7FOp3gQ> (дата звернення 22.05.2026 р.).
6. Калач Г., Шпак О., Кругляно А. Штучний інтелект в управлінні: автоматизація процесів та прийняття рішень // Соціальний розвиток: економіко-правові проблеми. – 2025. – № 5
7. BACnet, Modbus, KNX and OPC UA – what is used when in modern building automation? Електронний ресурс. URL: <https://zaphire.eu/bacnet-modbus-knx-og-opc-ua/> (дата звернення 22.05.2026 р.).
8. What are smart buildings? Електронний ресурс. URL:

<https://www.ibm.com/think/topics/smart-buildings> (дата звернення 22.05.2026 р.).

9. Баширов Е. Т., Бредіхін В. М. Підсистема обробки та валідації даних у системі підтримки рішень з енергоменеджменту / матеріали III (IX) Міжнародної науково-практичної конференції здобувачів вищої освіти і молодих учених «Інформаційні технології: теорія і практика», С. 430-432.

10. Oracle Електронний ресурс. URL: <https://www.oracle.com/ua/> (дата звернення 22.05.2026 р.).

11. García S. et al. Data preprocessing in data mining // Knowledge-Based Systems. – 2021. – Vol. 215.

12. Sarker I. H. Machine learning for intelligent systems // Journal of Big Data. – 2021. – Vol. 8.

13. Mahdavinejad M. S. et al. Machine learning for IoT data analytics // Digital Communications and Networks. – 2021.

14. Minoli D. Building the Internet of Things with IPv6 and MIPv6. – Wiley, 2021. – 400 p.

15. Da Xu L., He W., Li S. Internet of Things in industries // IEEE Transactions on Industrial Informatics. – 2021.

16. Vermesan O., Bacquet J. Internet of Things – The Call of the Edge. – River Publishers, 2022.

17. Al-Fuqaha A. et al. Internet of Things: A survey // IEEE Communications Surveys & Tutorials. – 2021.

18. Atzori L., Iera A., Morabito G. The Internet of Things: Vision and applications // Computer Networks. – 2021.

19. Gubbi J. et al. IoT architecture and applications // Future Generation Computer Systems. – 2022.

20. Wang S. et al. Smart building energy management systems: A review // Renewable and Sustainable Energy Reviews. – 2022.

21. Сидоренко В. О. Методи обробки даних у системах підтримки прийняття рішень // Інформаційні технології та комп'ютерна інженерія. –

2021. – № 3.

22. Робота з нестабільними даними у прогнозуванні часових рядів: стратегії та алгоритми Електронний ресурс. URL: <https://ua.linkedin.com/pulse/dealing-erratic-data-time-series-forecasting-ayush-chauhan?tl=uk> (дата звернення 25.05.2026 р.).

23. Zhong W., Suo Q., Jia X., Zhang A. And Su L. «Heterogeneous Spatio-Temporal Graph Convolution Network for Traffic Forecasting with Missing Values,» in 2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS), DC, USA, 2021 pp. 707–717. <https://doi.org/10.1109/ICDCS51616.2021.00073>

24. ARIMA Електронний ресурс. URL: <https://ru.wikipedia.org/wiki/ARIMA> (дата звернення 25.05.2026 р.).

25. Руденко О. Г., Безсонов О. О., Романюк О. С. Про один алгоритм навчання нейронної мережі в задачі прогнозування часових рядів // Біоніка інтелекту. 2018. № 1(90). С. 79—83.

26. MUSCLE Електронний ресурс. URL: <https://uk.wikipedia.org/wiki/MUSCLE> (дата звернення 27.05.2026 р.).

27. ESPRIT Електронний ресурс. URL: <https://uk.wikipedia.org/wiki/ESPRIT> (дата звернення 27.05.2026 р.).

28. Fahad Radhi, Denes Csala A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based TimeSeries Approac Inventions 2022, 7(4), 94; <https://doi.org/10.3390/inventions7040094>

29. Тест Дики-Фуллера (ADF) Електронний ресурс. URL: <https://mlaccessible.medium.com/%D1%82%D0%B5%D1%81%D1%82-%D0%B4%D0%B8%D0%BA%D0%B8-%D1%84%D1%83%D0%BB%D0%BB%D0%B5%D1%80%D0%B0-adf-a378b2c37143> (дата звернення 29.05.2026 р.).

30. UML Електронний ресурс. URL: <https://ru.wikipedia.org/wiki/UML> (дата звернення 29.05.2026 р.).

31. Python Електронний ресурс. URL: <https://www.python.org> (дата звернення 29.05.2026 р.).
32. A Python library to read/write Excel 2010 xlsx/xlsm files [Електронний ресурс]. URL: <https://openpyxl.readthedocs.io/en/stable/> (дата звернення 29.05.2026 р.).
33. Pandas Електронний ресурс. URL: <https://pandas.pydata.org/> (дата звернення 29.05.2026).
34. Anomaly Detection Toolkit, User Guide [Електронний ресурс]. URL: <https://arundo-adtk.readthedocshosted.com/en/stable/userguide> (дата звернення 29.05.2026 р.).
35. Statsmodels Електронний ресурс. URL: <https://www.statsmodels.org/stable/index.html> (дата звернення 02.06.2026 р.).
36. Закон України «Про охорону праці» / Офіційний сайт Верховної Ради України. – Режим доступу: <https://zakon.rada.gov.ua/go/2694-12> (дата звернення 02.06.2026 р.).
37. Охорона праці в офісі. Вимоги до робочого місця офісного працівника. – Режим доступу: <https://gc.ua/uk/oxorona-praci-v-ofisi-vimogi-do-robochogo-miscya-ofisnogo-pracivnika> (дата звернення 05.06.2026 р.).
38. Gogoi M., Kalita M. (2025) Assessment of Occupational Health Hazards of the Computer Users. J Ergonomics. 15:431.
39. Guidance on risk assessment at work. – Режим доступу: <https://osha.europa.eu/en/legislation/guidelines/guidance-risk-assessment-work> (дата звернення 05.06.2026 р.).
40. OSHA Electrical Safety Basics. – Режим доступу: <https://www.osha.gov/electrical> (дата звернення 07.06.2026 р.).
41. Computer Workplace Ergonomics. – Режим доступу: <https://cwa-union.org/national-issues/health-and-safety/health-and-safety-factsheets/computer-workplace-ergonomics> (дата звернення 07.06.2026 р.).

ДОДАТКИ

ДОДАТОК А.

Специфікація варіантів використання

Специфікація проєктованих варіантів використання наведено у таблицях А1 – А4.

Таблиця А1 – Специфікація варіанта використання «Надіслати сирі дані»

| |
|---|
| Прецедент: Надіслати сирі дані. |
| ID: 1. |
| Анотація: Надсилання даних для перетворення на препроцесор |
| Головні актори: External system (Зовнішня система) |
| Другорядні актори: Ні |
| Передумови: Ні |
| Основний потік: 1. Потік ініціюється, коли система надсилає дані в препроцесор; 2. Препроцесор запускає перетворення даних; |
| Постумови: Ні. |
| Альтернативні потоки: Якщо формат файлу не підтримується або дані не консистентні, система повертає повідомлення про помилку. |

Таблиця А2 – Специфікація варіанта використання «Отримати навчальну вибірку»

| |
|--|
| Прецедент: Отримати навчальну вибірку. |
| ID: 2. |
| Анотація: У зовнішню стему повертаються дані після відновлення нормалізації та видалення викидів. |
| Головні актори: External system (Зовнішня система) |
| Другорядні актори: Ні |
| Передумови: Дані мають пройти перетворення. |
| Основний потік: 1. Потік ініціюється, коли препроцесор завершить перетворення даних; 2. Уняшня система отримує дані у вигляді csv-файлу. |
| Постумови: Ні. |
| Альтернативні потоки: Ні |

Таблиця А3 – Специфікація варіанта використання «Отримати відновлені дані»

| |
|---|
| Прецедент: Отримати відновлені дані. |
| ID: 3. |
| Анотація: Зовнішня система зберігає в csv файл проміжні дані. |
| Головні актори: External system (Зовнішня система) |
| Другорядні актори: Ні |
| Передумови: Дані повинні пройти етап відновлення та очищення викидів без нормалізації. |
| Основний потік: 1. Потік ініціюється, коли завершився етап видалення викидів та відновлення; 2. Зовнішня система отримує csv-файл з даними. |
| Постумови: Ні. |
| Альтернативні потоки: Дані не записалися у файл; |

Таблиця А4 – Специфікація варіанта використання «Перетворити дані»

| |
|---|
| Прецедент: Перетворення даних. |
| ID: 4. |
| Дані надійшли на алгоритми препроцесора. |
| Головні актори: Ні |
| Другорядні актори: Ні |
| Передумови: Дані повинні бути консистентними після видалення файлу. |
| Основний потік: Потік ініціюється, коли з парсер приходять дані на алгоритм відновлення. |
| Постумови: Ні. |
| Альтернативні потоки: Ні |