

## РЕГРЕСІЙНІ МОДЕЛІ ДЛЯ АНАЛІЗУ ДАНИХ ІНТЕЛЕКТУАЛЬНОГО БУДИНКУ

О. Сінькевич

*Львівський національний університет імені Івана Франка,  
вул. Драгоманова, 50, 79005 Львів, Україна  
[oleh.sinkevych@lnu.edu.ua](mailto:oleh.sinkevych@lnu.edu.ua)*

У статті розглядається регресійно-кореляційне моделювання даних інтелектуального будинку з метою побудови як прогностичних моделей, так і дослідження зв'язків між ними. В якості даних для такого моделювання було обрано дві окремі групи: набори вимірювань температур та споживання енергії, отриманих в межах проекту REFIT Smart Homes та розподіли температур, виміряних у лабораторії інтелектуальних автономних систем факультету електроніки та комп'ютерних технологій Львівського національного університету імені Івана Франка за період з 1 лютого 2021 року по 1 вересня 2021 року. Розглянутий та апробований підхід до попередньої обробки кліматичних показників інтелектуального будинку, який передбачає застосування методу *STL*-декомпозиції. Побудова та дослідження регресійно-кореляційних моделей були виконані для кількох комбінацій даних: а) внутрішніх і зовнішніх температур; б) споживання газу і температур на обігрівальних елементах та в) внутрішніх, зовнішніх температур і споживання газу для обігріву. Результати такого моделювання дають змогу кількісно оцінити теплову ефективність інтелектуального будинку в контексті розподілів температур та споживання енергії обігріву, а положення регресійних прямих – неявним чином аналізувати теплозберігаючі властивості.

*Ключові слова:* регресійне моделювання, інтелектуальний будинок, XGBoost, інтелектуальний аналіз даних.

### Вступ та огляд проблематики

Розроблення систем інтелектуального будинку (ІБ) з використанням сучасних підходів на базі статистичної обробки даних та машинного навчання передбачає опрацювання даних та встановлення кількісних взаємозв'язків між ними. Оскільки більшість задач в контексті ІБ стосуються саме проблем забезпечення енергоефективності, оптимізації енерговитрат для опалення та підтримки комфортних температурних умов, необхідним етапом проектування є побудова математичних або статистичних моделей теплової поведінки, які пов'язують дані споживання енергії обігріву, температур, вологості тощо. Такому моделюванню передуює процес обробки та очищення відповідних сенсорних даних ІБ.

Необроблені дані від сенсорів є складними, характеризуються значним об'ємом та динамічними змінами, а відтак часто вимагають усунення шумів, зокрема Вейвлет-перетворенням [1]. Обробка пропущених значень відносно визначеного часового кроку має вагомий вплив, оскільки алгоритми аналізу даних переважно передбачають їх повноту та цільність, а наявність пропусків зумовлює низьку точність відповідних моделей.

Існує низка підходів до вирішення цієї проблеми, що ґрунтуються як на машинному навчанні, так і класичних чисельних методах [2-4]. В реальному середовищі різні сенсори IoT системи є розподіленими, а помилки у їх налаштуваннях та апаратні збої можуть продукувати отримання помилкових вимірювань, наприклад надто великих, малих чи некоректних значень. Такі аномалії можуть бути як точковими, контекстними (у часових рядах), так і у вигляді інтервалів підпослідовностей періодичного характеру. Розв'язування задачі виявлення таких аномалій здійснюється статистичними чи методами машинного навчання [5].

Регресійний та кореляційний аналіз разом з описовою статистикою та розвідувальним аналізом даних дає розуміння розподілів та взаємозв'язків кліматичних даних ІБ. Так у роботах [6, 7] здійснений регресійно-кореляційний аналіз температурних та даних споживання енергії та побудовані лінійні прогностичні регресійні моделі. У роботі [8] поряд з вичерпним розвідувальним аналізом побудовані регресійні залежності між середньодобовою тривалістю обігріву та середньодобовою температурою зовнішнього повітря, а у роботі [9] розглянута регресійна модель для аналізу даних про енергоспоживання двох ІБ. З метою покращення результатів лінійних регресійних моделей, у [10] авторами пропонується нова модель ансамблевого прогнозування на основі регресії з вбудованою автоматизованою оптимізацією параметрів для прогнозування споживання електроенергії; в [11] в якості регресійної моделі була використана LSTM-нейромережа. Оскільки комбінація багатьох сенсорів в ІБ може вносити різний вклад у точність моделей, у статті [12] регресійний підхід на базі MIDFEL LASSO регресії розвинутий для визначення найвпливовішого датчика в кожному сегменті горизонту прогнозу.

Реальні дані ІБ часто доволі складно піддаються адекватного моделюванню лінійними моделями, тому виникає необхідність у застосуванні таких підходів, як регресія випадковим лісом, або градієнтним підсиленням [13].

Визначені статистичні параметри, наприклад, коефіцієнт детермінації у моделі між зовнішніми та внутрішніми температурами без обігріву може неявно та наближено оцінювати теплову інерційність ІБ і властивість зберігати тепло будівлею. Регресійна модель, наведена в [14], дозволяє визначати теплові навантаження у будинку на основі фізичних властивостей будівельних компонент.

В цілому, регресійний аналіз є вагомим і ключовою компонентною при створенні цифрових аналітичних рішень для ІБ.

### **Опис та характеристика даних**

Оскільки накопичення достатньо великої кількості власних сенсорних даних, що вичерпно описують поведінку ІБ у розумінні динаміки розподілу кімнатної температури у комплексі зі споживанням енергії обігріву, не є завжди можливим внаслідок відсутності усіх необхідних сенсорних приладів (наприклад, для вимірювання точної кількості енергії споживання обігріву) чи можливості їх встановлення, у даній роботі було використані сторонні відкриті дані. *REFIT* – це бази даних різноманітних вимірювань, проведених у двадцяти будинках в межах наукового проекту *REFIT: Smart Homes and Energy Demand Reduction* [15].

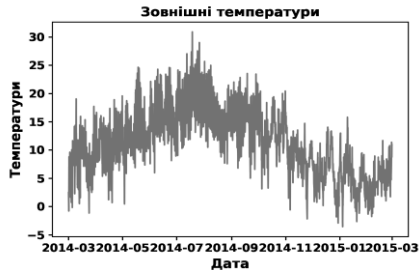


Рис. 1.1. Зовнішні температури



Рис. 1.2. Середні кімнатні температури



Рис. 1.3. Середні температури на обігрівальних елементах



Рис. 1.4. Сумарне споживання газу



Рис. 2.5. Температури у лабораторії

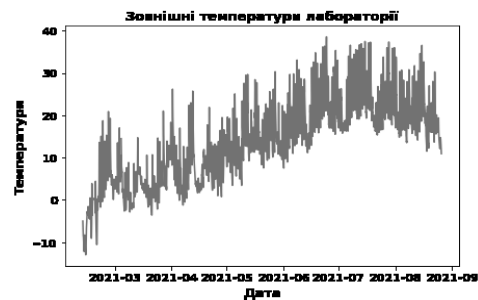


Рис. 2.6. Зовнішні температури поза лабораторією

У роботі з метою побудови і дослідження статистичних моделей було розглянуто динаміку на протязі одного року (з 1 березня 2014 р. по 1 березня 2015 р.) таких параметрів: а) зовнішні температури повітря; б) середні кімнатні температури у будинках; в) температури на обігрівальних елементах у кімнатах; г) значення споживання газу у кВт. Вимірювання цих параметрів проводились через одногодинний часовий інтервал. На рис. 1.1-1.4 наведено графіки для даних з бази *REFIT*, які відповідають першому будинку.

В свою чергу, на рис. 1.5-1.6 зображені температурні дані, виміряні у лабораторії інтелектуальних автономних систем факультету електроніки та комп'ютерних технологій Львівського національного університету імені Івана Франка за період з 1 лютого 2021 року по 1 вересня 2021 року. Ці дані були накопичені окремою апаратною системою на базі мікрокомп'ютера *Raspberry Pi3 Model B* та температурних датчиків *DS18B20*.

### Попередня обробка даних

Процес обробки сенсорних даних та їх статистичний аналіз необхідно здійснювати після етапу очищення, який передбачає: а) виявлення та заповнення пропущених значень й інтервалів; б) виявлення точкових викидів, аномалій чи аномальних підпоследовностей. Відомі підходи до проблеми, в яких для заповнення точкових пропусків використовується підхід вставки попереднього значення, а для інтервалу пропусків — лінійна інтерполяція чи ковзне середнє. Недоліками таких підходів є потреба у підборі оптимальної ширини ковзного вікна, що залежить від кількості пропусків, а при застосуванні лінійної інтерполяції — не враховується часові закономірності у даних на попередніх відрізках часу.

Для виявлення точкових викидів та аномалій у сенсорних даних застосуємо підхід на базі *STL*-декомпозиції. Для демонстрації ефективності цих методів додамо 1.5% аномальних точок (викидів) шляхом збільшення на деяку величину  $\delta_i = \pm 10$  випадкових температур у лабораторії за період з 1 березня по 1 квітня 2021 року (рис. 3). На рис. 3 (1) зображені додані аномалії, а на рис. 3 (2) результати адитивної -декомпозиції — графік виділеного тренду  $t_n$ , сезонна компонента  $s_n$  та різниці  $r_i$ , які визначаються як  $r_i = y_i - t_n - s_n$ , де  $y_i$  — температурний ряд. За розподілом різниць  $r_i$  та визначеним пороговим значенням  $\tau \in r_i$  можна виділяти аномальні значення, проте оцінка порогу потребує візуалізації експериментів.

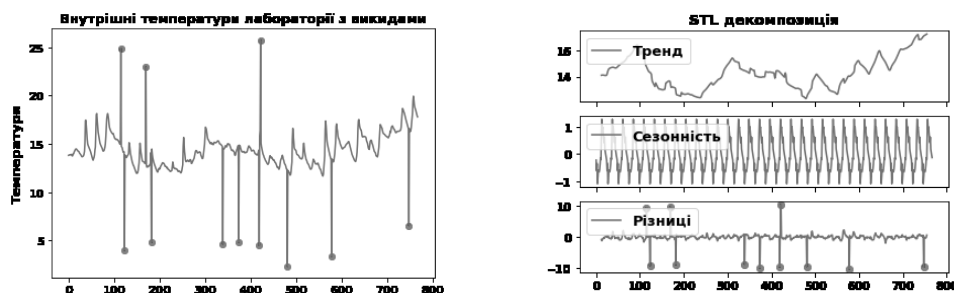


Рис. 3. Виявлення аномалій за *STL*-декомпозицією

### Візуалізація розподілів даних

Кліматичні дані ІБ, а також зв'язані з ними вимірювання споживання енергії обігріву зумовлюють необхідність виявлення структури взаємозалежностей, які відіграють важливу роль у побудові прогностичних моделей. Автоматизація такого процесу дозволяє надавати користувачу актуальну інформацію як про кліматичний стан ІБ, так і доповнювати загальну статистику щодо використання пристроїв обігріву, або охолодження приміщень. З цією метою розглянемо процес побудови статистичних регресійних моделей даних на прикладі *REFIT* та розподілів температур у лабораторії.

На рис. 4 показані діаграми розмаху ІБ *REFIT*: зовнішні температури, кімнатні температури, температури на обігрівальних елементах та споживання газу. Як можна побачити, зовнішні температури демонструють передбачувану річну поведінку з виділеним піком на теплу пору року, динаміку якої можна використовувати у статистичному моделюванні. Розподіл середніх кімнатних температур свідчить про наявність позитивної кореляції з зовнішніми температурами, хоча для більш точного аналізу можна розгляда-

ти такі залежності покімнатно. Оскільки протягом літнього періоду обігрів є відсутній, регресійна модель між незалежними зовнішніми та залежними внутрішніми температурами дозволяє встановити наближену оцінку теплового характеру будинку в сенсі його теплозберігаючих властивостей. Розподіл значень температур на обігрівальних елементах протягом теплового сезону дає підстави знехтувати ними, оскільки температури на батареях наближено рівні з внутрішньокімнатними. Проте, протягом холодної пори року такі виміри можна використовувати у регресійних моделях з урахуванням споживання газу.

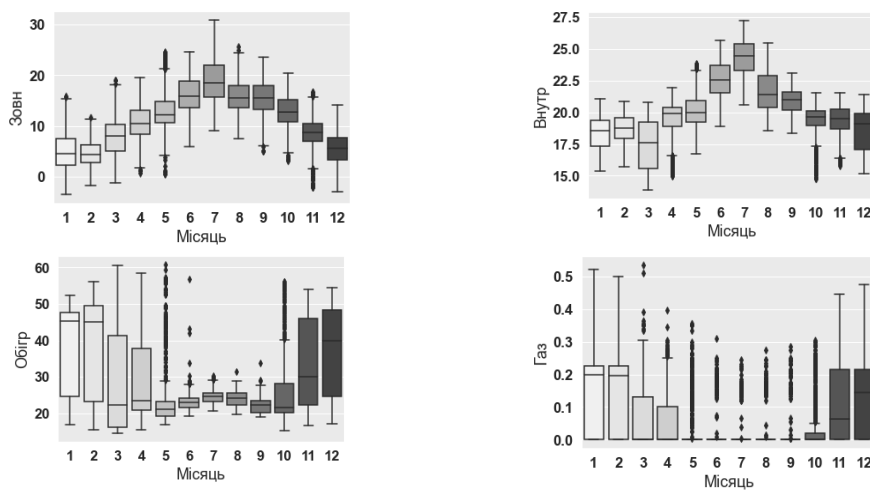


Рис. 4. Діаграми розмаху REFIT

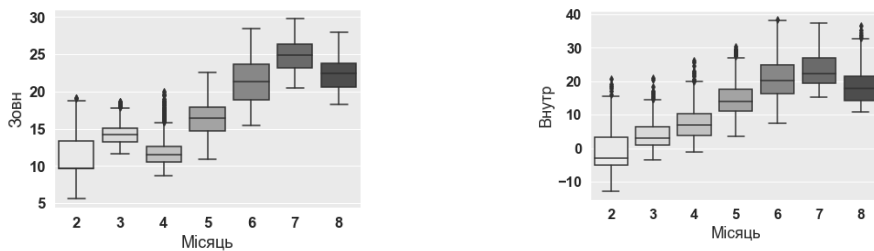


Рис. 5. Діаграми розмаху даних лабораторії

Розподіл значень температур на обігрівальних елементах протягом теплового сезону дає підстави знехтувати ними, оскільки температури на батареях наближено рівні з внутрішньокімнатними. Проте, протягом холодної пори року такі виміри можна використовувати у регресійних моделях з урахуванням споживання газу. В свою чергу, місячні розподіли використання газу протягом літнього періоду надають інформацію для задачі дезагрегації (визначення конкретних споживачів газу на базі сумарних значень та виявлених ознак), що є корисним для аналізу економії та оцінки теплової поведінки ІБ.

На рис. 5 зображені місячні статистики температур, виміряних у лабораторії. Динаміка даних протягом літнього періоду також може використовуватися для оцінки теплової інерції будівлі.

Відтак, розглянемо процес побудови та відбору регресійних моделей для встановлення залежностей між даними.

### Регресійні моделі для внутрішньої та зовнішньої температур

Розглянемо побудову лінійної регресійної моделі виду  $t_{ind} = f(t_{out}; \beta) + \varepsilon_t$ , де  $t_{out}$  – зовнішні температури протягом літнього періоду,  $t_{ind}$  – середні кімнатні температури,  $f(\cdot)$  – регресійна функція,  $\beta$  – параметри моделі, а  $\varepsilon_t$  – випадкова похибка. Така регресійна модель не лише дозволяє встановити залежність між температурами, а й наближено оцінити теплову інерцію будинку без наявності обігріву.

Здійснені чисельні експерименти за вищенаведеним алгоритмом, в результаті була отримана проста лінійна регресійна модель виду (рис. 6):

$$\ln(t_{ind}) = 2.84 + 0.014 \cdot t_{out} + \varepsilon_t, \quad (1)$$

де розподіл похибок  $\varepsilon_t$  показаний на гістограмі рис. 6 (2), а коефіцієнт детермінації  $R_{adj}^2 = 0,6$ . Логарифмування залежної змінної було проведено задля виправлення гетероскедастичності похибок  $\varepsilon_t$ , виявлених під час експериментів.

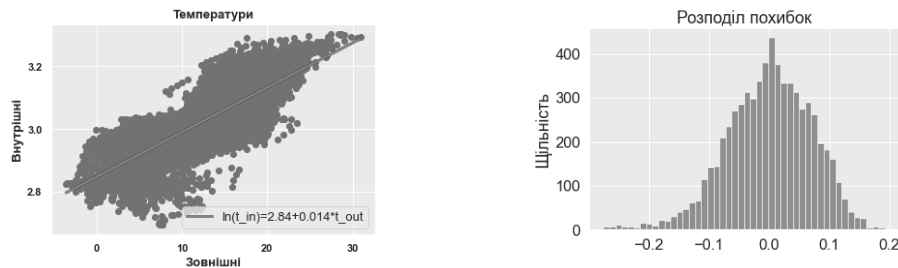


Рис. 6. Регресійна пряма для температур *REFIT* та розподіл похибок

Значення  $R_{adj}^2$  свідчить про те, що залежна величина  $t_{ind}$  кімнатних температур не в повній мірі залежить від зовнішньої температури  $t_{out}$ , тобто присутні інші фактори (явища конвекції тощо). Зокрема, величина  $R_{adj}^2$  залежить від теплофізичних параметрів будинку. Очевидно, чим менший кут нахилу регресійної прямої, тим більшою є теплова інерція будинку.

Для температур, виміряних у лабораторії, лінійна регресійна модель матиме вигляд (рис. 7 (1)):

$$\ln(t_{ind}) = 2.76 + 0.017 \cdot t_{out} + \varepsilon_t, \quad (2)$$

з розподілом похибок (рис. 7 (2)) та коефіцієнтом детермінації  $R_{adj}^2 = 0.6$ . Тут похибки  $\varepsilon_t$  не є нормально розподіленими, що може пояснюватися наявністю аномалій та викидів, не виявлених під час очищення даних. Позаяк, у моделі забезпечена скедастичність, лінійність та відсутність автокореляцій.

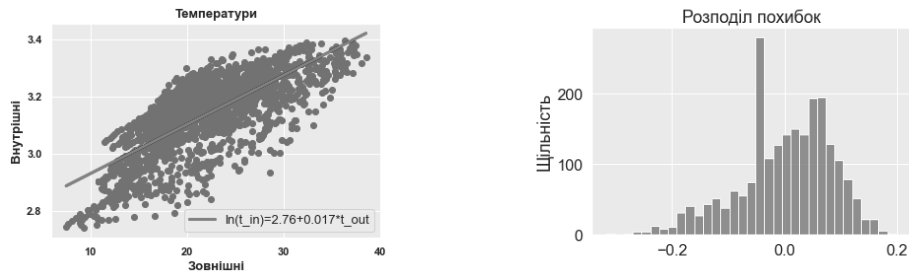
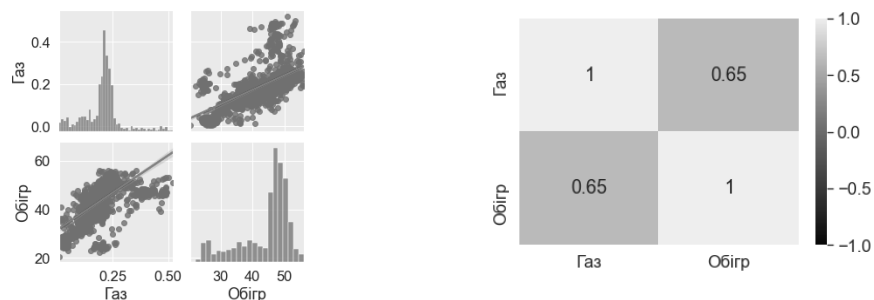


Рис. 7. Регресійна пряма для температур у лабораторії та розподіл похибок

### Регресійна модель для споживання газу та температури обігрівальних елементів

Наступна регресійна модель – моделювання взаємозв'язку між сумарним споживанням газу *gas* та температур на обігрівальних елементах *rad* протягом зимового періоду. Ті пари значень, де споживання газу рівне нулю, були видалені з вхідного масиву, оскільки вони не є інформативними для моделювання.

Рис. 8. Діаграма розсіювання даних споживання газу та температур на батареях *REFIT*

Як можна побачити з діаграми розсіювання рис. 8, споживання газу та температури не демонструють чіткої лінійної взаємозалежності, тому в даному випадку можна використати непараметричний регресор на основі градієнтного підсилювання. Також, таку залежність можна моделювати поліноміальною лінійною регресією, що потребує оптимального підбору степенів поліному.

Кореляційна матриця (коефіцієнти Пірсона) показує позитивну кореляцію між цими змінними (збільшення кількості споживання газу відповідає збільшенню температури на батареї).

На рис. 9 продемонстрований результат побудови регресійних моделей на базі полінома другого ступеня та градієнтного підсилення з використанням бібліотеки *XGBoost*.

Лінійна регресійна модель на базі поліному другого степеня вигляд (рис. 9(1)):

$$rad = 22.1 + 164.9 \cdot gas - 237.8 \cdot gas^2 + \varepsilon, \quad (3)$$

де коефіцієнт детермінації  $R_{adj}^2 = 0.6012$ , а значення середньої абсолютної відсоткової помилки ( $MAPE$ ) – 0.087.

Регресійна модель на основі градієнтного підсилення (рис. 9 (2)) має такі параметри: максимальна глибина побудованого дерева – 5, темп навчання – 0.3, значення середньої абсолютної відсоткової помилки ( $MAPE$ ) – 0.079.

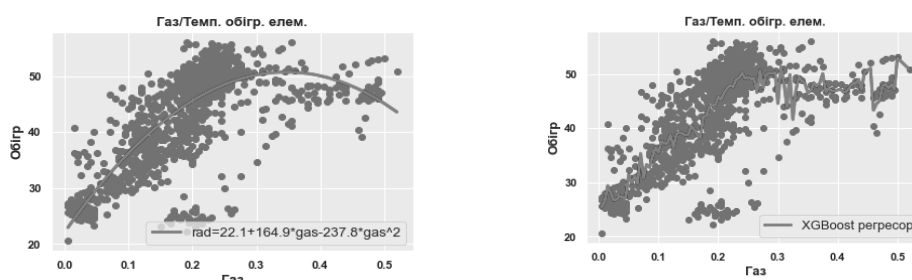


Рис. 9. Регресійні криві: поліноміальна та на основі градієнтного підсилення

Грунтуючись на показнику  $MAPE$ , регресійна модель градієнтного підсилення є точнішою та не вимагає поліноміальної трансформації як у випадку простої лінійної регресії. Позаяк, регресор на основі градієнтного підсилення містить значну кількість гіперпараметрів, які необхідно додатково оптимізувати. Дану модель можна використовуватися для прогнозування температури на батареях за відомим сумарним споживанням газу у будинку, проте її інтерпретація, на відміну від лінійної регресії, є складною, що є недоліком.

**Регресійна модель для внутрішньої, зовнішньої температури та споживанням газу**

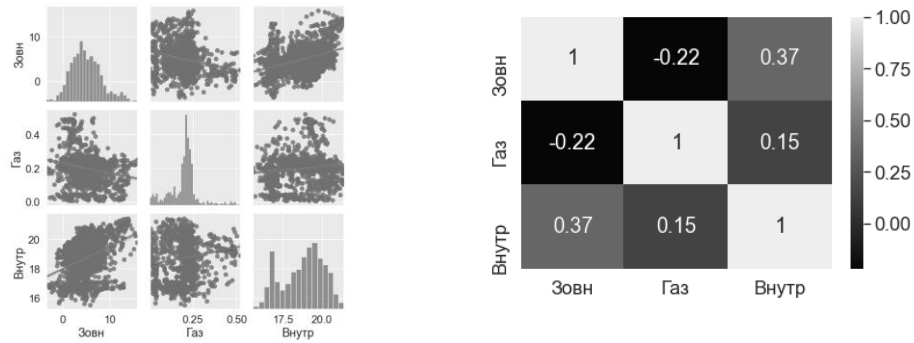
Розглянемо регресійну модель виду  $t_{ind} = f(t_{out}, gas; \beta) + \epsilon$ , що описує зв'язок між незалежними параметрами споживання газу  $gas$ , зовнішніми температурами  $t_{out}$  та залежними середніми внутрішніми температурами  $t_{ind}$  протягом зимового періоду. Попередньо триплети даних з нульовим споживанням газу були виключені з аналізу. Оскільки у моделі присутні дві змінні, їх можна протестувати (оцінка коефіцієнту інфляції дисперсії) на наявність мультиколінеарності (табл. 1). Значення цього коефіцієнту є менші за 10, що говорить про відсутність мультиколінеарності між сумарним споживанням газу та зовнішніми температурами.

На рис. 10 наведені діаграма розсіювання та кореляційна матриця, що свідчать про незначні кореляції та відсутність чітко вираженої лінійної залежності між даними.

Табл. 1. Оцінка коефіцієнту інфляції дисперсії

Змінна	Коефіцієнт інфляції дисперсії
$t_{out}$	3.99
$gas$	7.55



Рис. 10. Діаграма розсіювання даних споживання газу та температур *REFIT*

Відтак, для побудови прогностичної регресійної моделі використаємо алгоритм градієнтного підсилення, дерево якого зображено на рис 11 (параметри  $f_0$  та  $f_1$  відповідають  $t_{out}$  та  $gas$ ), а значення середньої абсолютної відсоткової помилки (*MAPE*) – 0.032.

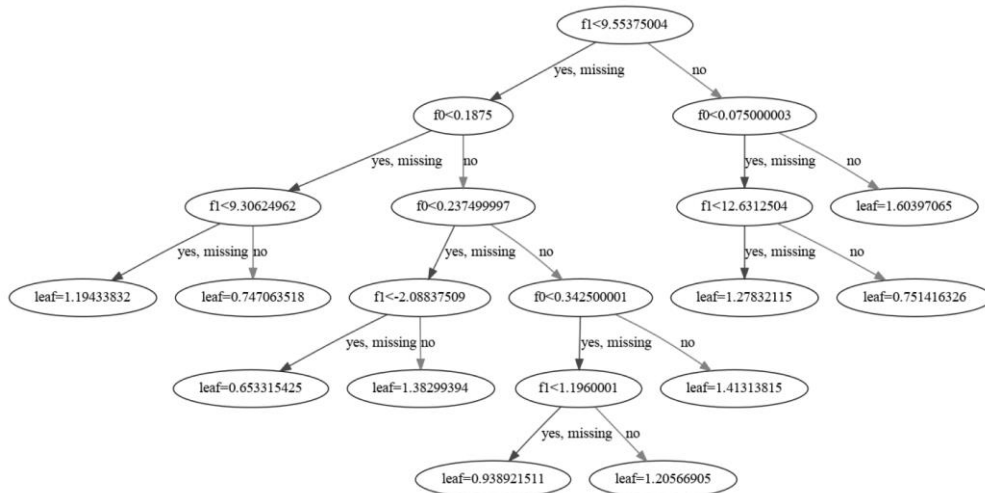


Рис. 11. Дерево моделі градієнтного підсилення

На відміну від регресійної моделі зовнішніх і внутрішніх температур протягом літа, зимовий період характеризується відсутністю вираженої прямої кореляції, що пов'язано з обігрівом приміщень. Врахування параметра сумарного споживання газу дозволяє більш точно змоделювати цей зв'язок, проте потребує окремого дослідження в контексті розв'язування задачі дезагрегації.

**Висновки**

У роботі побудовані та розглянуті статистичні моделі для аналізу та обробки сенсорних даних температур та споживання газу в ІБ. Описаний процес підготовки та попередньої обробки даних (виявлення викидів та аномальних показників) з використанням методу *STL*-декомпозиції. Розроблені регресійні моделі для:

а) температурних даних протягом літнього періоду за відсутності обігріву приміщень, на основі чого можна 1) наближено оцінити теплозберігаючі властивості кімнати/будинку шляхом аналізу кута нахилу регресійної прямої та величини коефіцієнту детермінації, що неявно вказують на здатність будинку підтримувати температурний режим під час втрачання тепла через зовнішні стіни внаслідок різниці температур; 2) здійснювати прогнозування внутрішніх температур за даними зовнішніми температурами;

б) даних споживання газу і температур на батареях протягом зимового періоду (лінійна модель з представленням незалежної змінної споживання газу у вигляді полінома другого степеня та модель на базі градієнтного підсилення), здійснено їх порівняння та виявлена перевага останньої, зокрема без обмежень, які типово накладаються на лінійну модель; така модель дозволяє пов'язати споживання газу та температури на батареях та у подальшому використовуватися для визначення ефективності обігріву;

в) даних зовнішніх та внутрішніх температур, а також споживання газу протягом зими, що дає змогу використовувати її для прогнозування внутрішніх температур за заданим споживанням газу і зовнішніми температурами; оскільки між даними немає чітко вираженої лінійної залежності, було використано метод на базі градієнтного підсилення, що повертає оптимальне регресійне дерево рішення.

Розроблені прогностичні регресійні моделі можуть використовуватися як для неявної оцінки теплозберігаючих характеристик ІБ, так і для статистичного аналізу теплової поведінки ІБ. Наприклад, значення коефіцієнтів детермінації, кутів нахилу регресійних прямих неявним чином свідчать як про ефективність опалення чи властивості ІБ до збереження температур, так і можуть доповнювати більш складні моделі для оптимізації функціонування ІБ.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Wavelet-Based filtration procedure for denoising the predicted CO2 waveforms in smart home within the internet of things / Jan Vanus [et al.] // *Sensors*. – 2020. – V. 20, № 3. – P. 620. – Retrieved from: <https://doi.org/10.3390/s20030620>.
- [2] Missing value imputation based on gaussian mixture model for the internet of things / Xiaobo Yan [et al.] // *Mathematical problems in engineering*. – 2015. – V. 2015. – P. 1–8. – Retrieved from: <https://doi.org/10.1155/2015/548605>.
- [3] A missing sensor data estimation algorithm based on temporal and spatial correlation / Zhipeng Gao [та ін.] // *International journal of distributed sensor networks*. – 2015. – V. 2015. – P. 1–10. Retrieved from: <http://dx.doi.org/10.1155/2015/435391>.
- [4] *Mary I. P. S.* Imputing the missing data in IoT based on the spatial and temporal correlation / I. Priya Stella Mary, L. Arockiam // 2017 IEEE international conference on current trends in advanced computing (ICCTAC), Bangalore, 2–3 Mar. 2017. – [Б. м.], 2017.
- [5] *Lin W.-C.* Deep learning for missing value imputation of continuous data and the effect of data discretization / Wei-Chao Lin, Chih-Fong Tsai, Jia Rong Zhong // *Knowledge-Based*

- systems. – 2022. – V. 239. – P. 108079. Retrieved from: <https://doi.org/10.1016/j.knosys.2021.108079>.
- [6] *Sinkevych O.* Statistical Analysis of the Thermal Parameters of Smart Homes / O. Sink-evych, L. Monastyrskii, B. Sokolovskyi // *Electronics and information technologies.* 2018. – Issue 10. – P. 99–108
- [7] *Oleh Sinkevych, Liubomyr Monastyrskiy, Bohdan Sokolovskyi.* Determination of regression parameters for the thermal and energy components of smart homes / *International Scientific and Practical Conference "Electronics and Information Technologies" (ELIT-2018).* A-92 A-95. 2018.
- [8] Heating behaviour in English homes: an assessment of indirect calculation methods / T. Kane [et al.] // *Energy and buildings.* – 2017. – V. 148. – P. 89–105. Retrieved from: <https://doi.org/10.1016/j.enbuild.2017.04.059>.
- [9] Linear regression analysis of energy consumption data for smart homes / Pheny Phemelo Moletsane [et al.] // *2018 41st international convention on information and communication technology, electronics and microelectronics (MIPRO), Opatija, 21–25 May. 2018 .* – [Б. м.], 2018. Retrieved from: <https://doi.org/10.23919/MIPRO.2018.8400075>.
- [10] Energy usage prediction for smart home with regression based ensemble model / Mohammad Shamsul Hoque [et al.] // *2020 8th international conference on information technology and multimedia (ICIMU), Selangor, Malaysia, 24–26 Aug. 2020 .* – [Б. м.], 2020. Retrieved from: <https://doi.org/10.1109/ICIMU49871.2020.9243578>.
- [11] *Shorfuzzaman M.* Predictive analytics of energy usage by iot-based smart home appliances for green urban development / Mohammad Shorfuzzaman, M. Shamim Hossain // *ACM transactions on internet technology.* – 2022. – V. 22, № 2. – P. 1–26. Retrieved from: <https://doi.org/10.1145/3426970>.
- [12] *Spencer B.* Forecasting temperature in a smart home with segmented linear regression / Bruce Spencer, Omar Alfandi, Feras Al-Obeidat // *Procedia computer science.* – 2019. – V. 155. – P. 511–518. Retrieved from: <https://doi.org/10.1016/j.procs.2019.08.071>.
- [13] *Ma X.* Prediction of outdoor air temperature and humidity using Xgboost / Xiaoming Ma, Cong Fang, Junping Ji // *IOP conference series: earth and environmental science.* – 2020. – V. 427. – P. 012013. Retrieved from: <https://doi.org/10.1088/1755-1315/427/1/012013>.
- [14] A regression-based framework to examine thermal loads of buildings / Mohammad K. Najjar [et al.] // *Journal of cleaner production.* – 2021. – V. 292. – P. 126021. Retrieved from: <https://doi.org/10.1016/j.jclepro.2021.126021>.
- [15] REFIT – REFIT: smart homes and energy demand reduction [Електронний ресурс] // REFIT – REFIT: Smart Homes and Energy Demand Reduction. – Retrieved from: <https://www.refitsmarthomes.org/>.

**REGRESSION MODELS FOR SMART HOME DATA ANALYSIS****O. Sinkevych**

*Ivan Franko National University of Lviv  
50 Drahomanova str., 79005, Lviv, Ukraine  
[oleh.sinkevych@lnu.edu.ua](mailto:oleh.sinkevych@lnu.edu.ua)*

The development of smart home systems using modern approaches based on statistical and machine learning involves processing data and establishing quantitative relationships between multiple measurements. Since most of the tasks in the context of smart home relate specifically to the problems of ensuring energy efficiency, optimizing energy consumption for heating and maintaining comfortable temperature conditions, a necessary stage of design is the development of mathematical or statistical models of thermal behavior that connect data on heating energy consumption, temperatures, humidity, etc. The paper presents the regression-correlation modeling of the smart home data with the aim of developing either predictive models and researching the relationships between corresponding data measurements. Two separate data groups were selected to conduct the research: sets of temperature and energy consumption measurements obtained within the REFIT Smart Homes project and temperature distributions measured in the laboratory of intelligent autonomous systems of the Faculty of Electronics and Computer Technologies of Ivan Franko Lviv National University the period from February 1, 2021 to September 1, 2021. The approach to the pre-processing of climatic parameters of the smart home, which involves the use of the STL-decomposition method, is considered and implemented. Development and research of the regression-correlation models were performed for several combinations of data: a) internal and external temperatures; b) gas consumption and temperatures on heating elements and c) internal and external temperatures and gas consumption used for heating. The developed predictive regression models can be used both for the implicit assessment of the heat-saving characteristics of the smart home and for the statistical analysis of the thermal behavior. For example, the values of the coefficients of determination, the angles of inclination of the regression lines implicitly determine the efficiency of heating process or the properties of the building to maintain temperatures, and can complement more complex models to optimize the functioning of the smart home.

*Key words:* regression modeling, smart home, XGBoost, data mining

*Стаття надійшла до редакції 29.11.2022.*

*Прийнята до друку 05.12.2022.*