

УДК 621.181.1:681.5

DOI <https://doi.org/10.32782/tnv-tech.2024.2.3>

## РОЗРОБКА МОДЕЛІ КОНДЕНСАЦІЙНОГО КОТЛА НА ОСНОВІ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПОБУДОВИ ЦИФРОВОГО ДВІЙНИКА

**Зінченко Д. Д.** – магістр кафедри автоматизації енергетичних процесів  
Національного технічного університету України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
ORCID ID: 0000-0001-5524-6938

**Новіков П. В.** – кандидат технічних наук,  
доцент кафедри автоматизації енергетичних процесів  
Національного технічного університету України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
ORCID ID: 0000-0002-2790-5809

**Волощук В. А.** – доктор технічних наук,  
професор кафедри автоматизації енергетичних процесів  
Національного технічного університету України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
ORCID ID: 0000-0003-0687-8968

**Штіфзон О. Й.** – старший викладач кафедри автоматизації  
енергетичних процесів  
Національного технічного університету України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
ORCID ID: 0000-0003-0011-4617

У статті розглянуто розроблення регресійної моделі конденсаційного котла як одного з компонентів цифрового двійника системи теплозабезпечення. ККД конденсаційного котла може сягати 96%, але він напряму залежить від споживача теплової енергії, а саме від температури зворотного теплоносія. Ефективна робота конденсаційного котла має місце лише при застосуванні низькотемпературних систем теплозабезпечення будинків із відповідними теплотехнічними характеристиками. Для забезпечення цих характеристик необхідно мати інтелектуальну систему керування. Система має відстежувати низку параметрів та обмежень, таких як температури теплоносія, витрата палива, склад димових газів, стан конструкційних матеріалів рівень конденсату тощо. Цифровий двійник конденсаційного котла має інтегрувати алгоритми керування, моделі та дані для побудови режимів експлуатації в реальному часі. Він забезпечить максимальну ефективність та ресурс обладнання при дотриманні всіх необхідних експлуатаційних обмежень. Метою дослідження є отримання статистичної моделі за допомогою методів машинного навчання, вибір та ранжування набору вхідних параметрів моделі, навчання моделей та аналіз їх точності. Регресійну модель розроблено на основі даних роботи конденсаційного котла під час експлуатації. Дані функціонування котлоагрегату, на основі яких розраховувалась модель, були зібрані під час його експлуатації, за період у 30 днів. Проведено процес навчання лінійної моделі регресії. Обрано для неї набір вхідних даних та розраховано коефіцієнти. Вхідними параметрами котла є температури вихідного теплоносія і димових газів. Для спрощення кінцевої моделі, було визначено набір вхідних параметрів що мають значущий вплив на вихідні параметри. Розраховано архітектуру регресивного дерева. Обрані моделі об'єднано ансамблевим методом «золотування». Проведено оцінку точності розрахунку моделей.

**Ключові слова:** конденсаційний котел, моделювання, регресійна модель, машинне навчання, цифровий двійник.

**Zinchenko D. D., Novikov P. V., Voloshchuk V. A., Shtifzon O. Y. Development of a condensing boiler model based on machine learning methods for building a digital twin**

The article deals with the development of a regression model of a condensing boiler as one of the components of a digital twin of the heat supply system. The efficiency of a condensing boiler can reach 96%, but it directly depends on the consumer of thermal energy, namely on the temperature of the return coolant. The efficient operation of the condensing boiler takes place only when using low-temperature heat supply systems for buildings with appropriate thermal characteristics. To ensure these characteristics, it is necessary to have an intelligent control system. The system must monitor a set of parameters and limitations, such as coolant temperatures, fuel consumption, composition of flue gases, condition of structural materials, condensate level, etc. The digital twin of the condensing boiler must integrate control algorithms, models and data to build operating modes in real time. It will ensure the maximum efficiency and resource of the equipment while observing all the necessary operational restrictions. The purpose of the research is to obtain a statistical model using machine learning methods, select and rank a set of model input parameters, train models and analyze their accuracy. The regression model was developed based on the operation data of the condensing boiler during operation. The operating data of the boiler unit, based on which the model was calculated, were collected during its operation, for a period of 30 days. The learning process of the linear regression model was carried out. A set of input data was selected for it and the coefficients were calculated. The input parameters of the boiler are the temperature of the return coolant and the consumption of natural gas. The initial parameters of the boiler are the temperature of the output coolant and flue gases. To simplify the final model, a set of input parameters that have a significant impact on the output parameters was determined. The architecture of the regression tree was calculated. The selected models are united by the ensemble method of "voting". The accuracy of model calculations was evaluated.

**Key words:** condensing boiler; modeling; regression model; machine learning; digital twin.

**Постановка проблеми.** Рівень зношеності теплової інфраструктури в Україні є дуже високим: 40% теплових пунктів і 15,8% тепломереж перебувають в аварійному стані [1]. Це викликає значні втрати енергії при транспортуванні теплоносія. Перехід на локальні автономні системи опалення з використанням конденсаційних котлів може допомогти подолати цю проблему шляхом децентралізації та модернізації.

Крім того, більшість житлового фонду в країні обладнана застарілими неефективними системами опалення та гарячого водопостачання. Щоб максимально реалізувати потенціал енергозбереження конденсаційних котлів, необхідна комплексна модернізація внутрішньобудинкових інженерних систем з встановленням низькотемпературних опалювальних приладів та ізоляцією трубопроводів.

Масштабне впровадження конденсаційних котлів в поєднанні з термомодернізацією будівель може суттєво знизити споживання енергоресурсів в житлово-комунальному секторі України, який традиційно є одним з найбільш енергозатратних. Це дозволить підвищити енергетичну безпеку країни, зменшити залежність від імпорту викопних видів палива та скоротити викиди парникових газів від теплоенергетичних об'єктів.

Для забезпечення ефективної роботи конденсаційного котла в оптимальному режимі необхідно розробити інтелектуальну систему керування на основі алгоритмів адаптивного управління та прогнозування. Система має відстежувати низку параметрів та обмежень, таких як температура теплоносія, витрата палива, склад димових газів, стан конструкційних матеріалів та рівень конденсату.

По-перше, алгоритм керування повинен регулювати температурний режим системи опалення, утримуючи її в низькотемпературному діапазоні для забезпечення максимальної конденсації водяної пари з димових газів. Це вимагає прогнозування теплового навантаження з урахуванням погодних умов, стану будівлі та графіків опалення.

По-друге, необхідно контролювати процес згоряння та регулювати подачу повітря і палива для мінімізації викидів шкідливих речовин. Алгоритми можуть використовувати дані з датчиків складу димових газів та адаптивно коригувати коефіцієнти надлишку повітря.

По-третє, система має відстежувати та прогнозувати стан вузлів і деталей котла з використанням методів машинного навчання та моделей деградації. Це дозволить завчасно виявляти потенційні несправності та планувати проактивне обслуговування.

Цифровий двійник конденсаційного котла має інтегрувати всі ці алгоритми керування, моделі та дані для побудови режимів експлуатації в реальному часі. Він забезпечить максимальну ефективність та ресурс обладнання при дотриманні всіх необхідних експлуатаційних обмежень.

**Аналіз останніх досліджень і публікацій.** З робіт [2–3] видно, що конденсаційні котли стають усе привабливішими для споживачів через свою високу ефективність (ККД > 90%) та постійне подорожчання ресурсів. Проте, досягнення високого показника ефективності при використанні конденсаційного котла можливо лише при експлуатації котлоагрегату при оптимальних умовах [11]. Такі умови може забезпечити лише правильно впроваджена система автоматизації. Система автоматизації складних об'єктів вимагає наявності моделі для синтезу регулятора. У роботах [4–9] описано способи отримання моделей у вигляді динамічних рівнянь. До переваг такого підходу можна віднести універсальність, тобто модель однаково описує усі об'єкти для яких була створена. Однак, це може бути і суттєвим недоліком, бо модель не відображає певні особливості конкретного об'єкта керування та може не враховувати зношення або певні особливості конкретного котла.

Найбільш широко застосовуваним і відомим методом отримання моделі об'єкту управління є апроксимація його перехідних характеристик за допомогою функцій передач [10, 11, 12]. Однак, його основним недоліком є необхідність проведення окремих експериментів для отримання розгінних характеристик. На працюючому обладнанні це не завжди є можливим.

Регресійні моделі, на відміну від класичних методів, пропонують більш гнучкий і адаптивний підхід до моделювання об'єктів управління. Вони можуть бути побудовані на основі вже наявних даних, зібраних під час нормальної експлуатації системи, що дозволяє уникнути проведення спеціальних експериментів. Це значно економить час і ресурси. Сучасні методи машинного навчання дозволяють регресійним моделям досягати високої точності прогнозування, що робить їх ефективною альтернативою класичним методам.

**Постановка завдання.** Метою дослідження є розробка регресійної моделі за допомогою методів машинного навчання, вибір та ранжування набору вхідних параметрів моделі, навчання моделей та аналіз їх точності.

#### **Виклад основного матеріалу дослідження.**

##### **Опис об'єкта управління**

Об'єктом управління є конденсаційний котел. Такі котли використовуються у системах опалення і гарячого водопостачання. Особливістю такого типу котлів є дуже високий ККД, що досягається за допомогою використання тепла конденсації продуктів згоряння палива. Розглянемо принцип роботи такого котла.

Спочатку приготована суміш газу з повітрям подається на пальник 1 через газохід 2. Для регулювання потужності використовується вентилятор 3 зі змінним числом обертів, яке контролюється блоком управління. Це дозволяє здійснювати

«модуляцію полум'я», тобто регулювання інтенсивності полум'я від 20% до 100% від номінальної потужності. Повітря для згоряння засмоктується вентилятором через сопло, яке пов'язане з газовим клапаном. Кількість газу, яка подається вентилятору, пропорційна кількості засмоктуваного повітря.

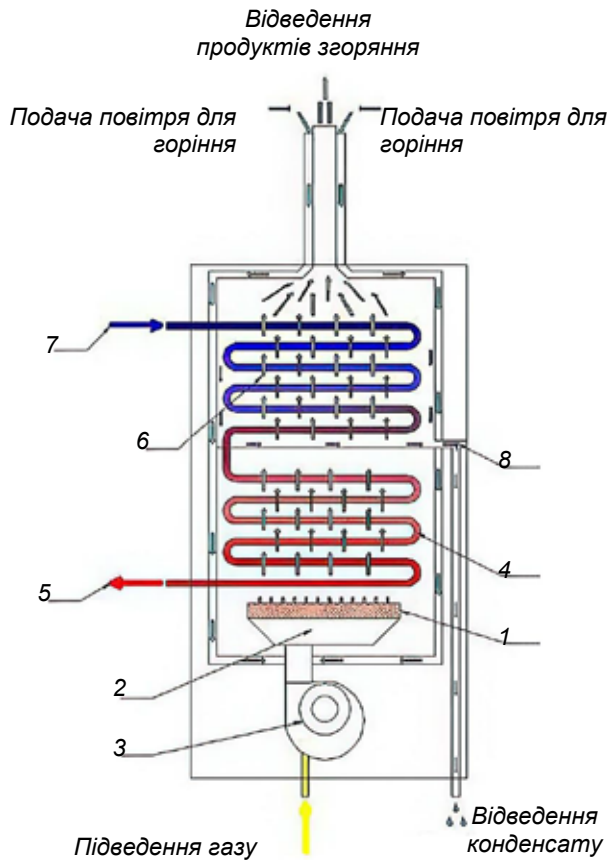


Рис. 1. Спрощена схема конденсаційного котла

Гарячі гази від спалювання газової суміші, проходячи через перший відсік теплообмінника 4, догрівають теплоносій, що звідти через трубопровід 5 потрапляють до споживача. Після цього, гази потрапляють у другий відсік теплообмінника 6, у якому відбувається первинне нагрівання зворотної води, що отримується від споживача через трубопровід 7. У цьому відсіку продукти згоряння охолоджуються нижче точки роси і водяна пара, що в них міститься, конденсується на стінках теплообмінника і відводиться через трубку 8. Цей процес конденсації дозволяє відновлювати додаткову теплову енергію, яка зазвичай втрачається при спалюванні палива.

При виході через димохід, залишкові продукти згоряння додатково підігрівають повітря, що подається в паливник, додатково збільшуючи ефективність процесу.

Для того, аби не виводити котел з режиму роботи, дані для отримання моделі були зібрані під час експлуатації котла. У такому випадку ідентифікація моделі

класичними методами є неможливою, бо неможливим є отримання розгінних характеристик. Тому, було створено дискретну модель котла за допомогою методів машинного навчання.

### Ідентифікація моделі

#### 1. Збір даних

Дані для створення моделі були зібрані під час експлуатації котла. Інтервал збору даних складав 1 с; період, за який зібрані дані – 30 днів. Параметри, що збирались:

- Температура вхідного теплоносія  $T_{вх}$
- Температура вихідного теплоносія  $T_{вих}$
- Температура димових газів  $T_{газ}$
- Витрата газу на пальник  $Q$

Остання шоста частина даних (5 днів) була залишена як тестова вибірка – навчання на ній не проводиться, тільки перевірка точності прогнозування моделями.

#### 2. Вибір вхідних та вихідних змінних

Вихідними параметрами котла є температури вихідного теплоносія і димових газів. Як цільовий параметр для моделі обрано зміни цих параметрів за 1 крок – це значно зменшить складність задачі для навчання за рахунок того, що зміни за крок є більш локальними – вони не враховують довгострокові тренди, відв'язуючись від абсолютного значення температури. Отже, цільовими (прогнозованими) параметрами є зміни між наступним і поточним кроком температури вихідного теплоносія  $\Delta T_{вих, t+1} = T_{вих, t+1} - T_{вих, t}$  і температури димових газів  $\Delta T_{газ, t+1} = T_{газ, t+1} - T_{газ, t}$ .

У ролі вхідних параметрів обрано як абсолютні значення температур вихідного  $T_{вих, t}$  та вхідного  $T_{вх, t}$  теплоносіїв, димових газів  $T_{газ, t}$  і витрати газу  $Q_t$ , так і їх зміни відносно минулого кроку:  $\Delta T_{вих, t}$ ,  $\Delta T_{вх, t}$ ,  $\Delta T_{газ, t}$ ,  $\Delta Q_t$  відповідно. Також, використано абсолютне значення витрати газу на наступний крок  $Q_{t+1}$  та її зміну  $\Delta Q_{t+1}$  як сигнал керування.

#### 3. Лінійна регресія

Лінійна регресія – один з найпростіших алгоритмів класичного машинного навчання [14]. У лінійній регресії залежна змінна є лінійною комбінацією параметрів, що в загальному вигляді вона формулою (1).

$$y = \beta_0 + \sum_{i=1}^n \beta_i \cdot x_i \quad (1)$$

де  $y$  – залежний параметр

$\beta_0$  – центрувальний коефіцієнт

$x_i$  – вільні параметри

$\beta_i$  – коефіцієнти вільних параметрів

Маючи тестові дані  $y_j$  й  $x_{ij}$ , для яких треба розрахувати коефіцієнти, складають систему рівнянь з (1) для кожного  $j$ .

Для визначення коефіцієнтів  $\beta_0, \beta_1, \dots, \beta_n$  застосовують метод найменших квадратів. Він полягає у спробі знайти функцію (2), яка мінімізує суму квадратів розбіжностей між правою та лівою сторонами системи.

$$S(\beta_0, \beta_1, \dots, \beta_n) = \sum_{j=1}^m \left( y_j - \left( \beta_0 + \sum_{i=1}^n \beta_i \cdot x_{ij} \right) \right)^2 \quad (2)$$

Мінімум обчислюють за допомогою розв'язання системи диференційних рівнянь через обчислення часткової похідної від (2) по кожному з коефіцієнтів і прирівняння їх до нуля (3).

$$\begin{cases} \frac{\partial S}{\partial \beta_0} = 0 \\ \frac{\partial S}{\partial \beta_1} = 0 \\ \dots \\ \frac{\partial S}{\partial \beta_n} = 0 \end{cases} \quad (3)$$

Розв'язок СЛАР (3) і є значення коефіцієнтів регресії.

За цим алгоритмом було розраховано коефіцієнти моделі для прогнозування вихідних параметрів. Точність розрахунку було оцінено за допомогою середньої квадратичної похибки. Для порівняння також було розраховано похибку фіктивного предиктора, який використовує як прогнозоване значення поточне значення прогнозованого параметру. Для цього застосуємо інструментарій мови python, а саме бібліотеки pandas, numpy та модулі linear\_model й metrics бібліотеки scikit-learn. Отримали модель такого вигляду:

$$\Delta T_{вих,t+1} = \beta_{T_{вих},0} + \sum_{i=1}^n \beta_{T_{вих},i} \cdot x_i, \quad (4)$$

$$\Delta T_{газ,t+1} = \beta_{T_{газ},0} + \sum_{i=1}^n \beta_{T_{газ},i} \cdot x_i, \quad (5)$$

$$x = \left[ T_{вих,t} \quad \Delta T_{вих,t} \quad T_{газ,t} \quad \Delta T_{газ,t} \quad Q_t \quad \Delta Q_t \quad Q_{t+1} \quad \Delta Q_{t+1} \right],$$

$$\beta_{T_{вих},0} = 0.00655,$$

$$\beta_{T_{вих}} = \begin{bmatrix} -0.00604 & 0.637 & 0.00553 & 0.137 & 0.000387 \\ 0.167 & -0.019 & 0.227 & 0.0437 & 0.0626 \end{bmatrix},$$

$$\beta_{T_{газ},0} = 0.0166,$$

$$\beta_{T_{газ}} = \begin{bmatrix} -0.015 & 0.256 & 0.0139 & 0.521 & 0.000862 \\ 0.203 & 0.422 & 2.05 & 0.484 & 0.907 \end{bmatrix}$$

Таблиця 1

### Порівняння похибок прогнозування лінійною регресією і фіктивним предиктором

Алгоритм	Середня квадратична похибка при прогнозуванні			
	$\Delta T_{вих,t+1}$		$\Delta T_{газ,t+1}$	
	навчальна вибірка	тестова вибірка	навчальна вибірка	тестова вибірка
Лінійна регресія	2.35e-5	4.21e-5	3.07e-4	4.29e-4
Фіктивний предиктор	1.61e-4	2.22e-4	7.31e-4	8.01e-4

#### 4. Ранжування набору вхідних параметрів

Для спрощення кінцевої моделі, було визначено набір вхідних параметрів що мають значущий вплив на вихідні параметри. Аби це зробити, потрібно дізнатись, на скільки сильний вплив мають значення окремих параметрів на значення

прогнозованої змінної. Для цього застосовано коефіцієнт кореляції Пірсона [15], який вимірює лінійну кореляцію між двома змінними. Він обчислюється формулою (4).

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6)$$

де  $n$  – розмір вибірки

$x_i, y_i$  – окремі точки вибірок

$\bar{x}, \bar{y}$  – середні значення вибірок

Коефіцієнти Пірсона було розраховано за допомогою функції `r_regression` модуля `feature_selection` бібліотеки `scikit-learn`.

Таблиця 2

**Коефіцієнт кореляції Пірсона для кожного з вхідних параметрів у відношенні до вихідних**

Вхідний параметр	Коефіцієнт Пірсона для	
	$\Delta T_{вих,t+1}$	$\Delta T_{саз,t+1}$
$T_{вих,t}$	-0.0023	-0.0026
$\Delta T_{вих,t}$	0.81	0.12
$T_{саз,t}$	-0.0043	-0.002
$\Delta T_{саз,t}$	0.042	0.016
$T_{Q,t}$	0.0016	-0.0021
$\Delta T_{Q,t}$	0.68	0.15
$Q_t$	0.0081	0.0021
$\Delta Q_t$	0.37	0.59
$Q_{t+1}$	0.0085	0.0042
$\Delta Q_{t+1}$	0.072	0.44

Для визначення набору значущих вхідних параметрів, було розраховано коефіцієнти лінійної регресії за допомогою методу найменших квадратів з різними кількостями вхідних параметрів: від 1 до 10 обираючи параметри з найбільшими абсолютними значеннями коефіцієнтів кореляції за табл. 1. Було оцінено точність розрахунку температур при кожному з наборів вхідних параметрів за допомогою середньої квадратичної похибки.

Як видно з рис. 2 і 3, після введення п'ятого параметру похибка практично перестала зменшуватись. Це число було обране як число значущих параметрів для подальших досліджень. Тоді, як для моделі  $\Delta T_{вих,t+1}$ , так і для  $\Delta T_{саз,t+1}$  вхідними параметрами будуть  $\Delta T_{вих,t}$ ,  $\Delta T_{саз,t}$ ,  $\Delta T_{Q,t}$  і  $\Delta Q_{t+1}$ . Отримали модель такого вигляду:

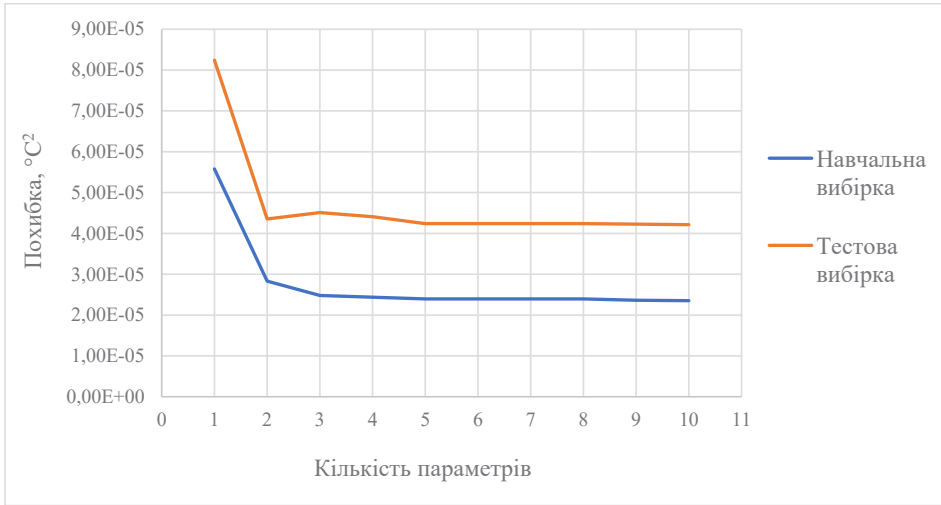


Рис. 2. Середня квадратична похибка розрахунку  $\Delta T_{вих,t+1}$  в залежності від кількості вхідних параметрів

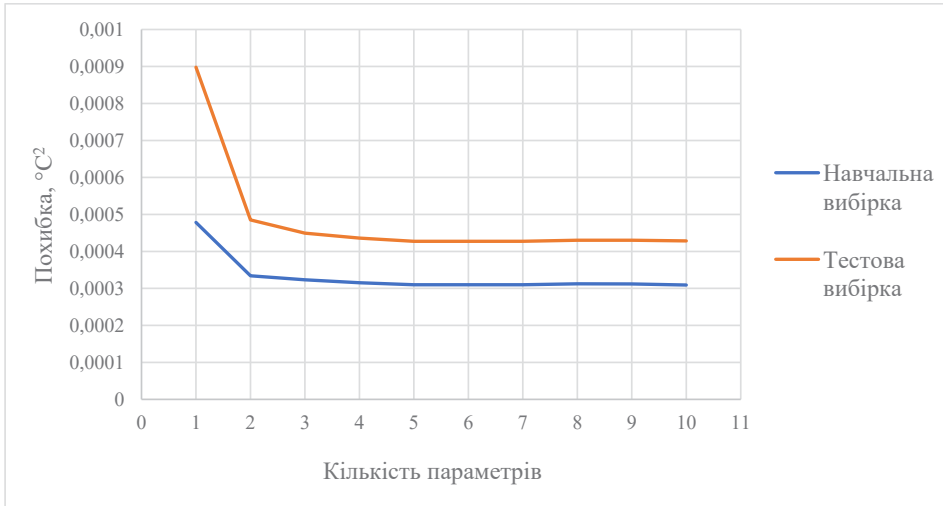


Рис. 3. Середня квадратична похибка розрахунку  $\Delta T_{газ,t+1}$  в залежності від кількості вхідних параметрів

$$\Delta T_{вих,t+1} = \beta_{T_{вих},0} + \sum_{i=1}^n \beta_{T_{вих},i} \cdot x_i, \quad (7)$$

$$\Delta T_{газ,t+1} = \beta_{T_{газ},0} + \sum_{i=1}^n \beta_{T_{газ},i} \cdot x_i, \quad (8)$$

$$x = [\Delta T_{вих,t} \quad \Delta T_{ох,t} \quad \Delta T_{газ,t} \quad \Delta Q_t \quad \Delta Q_{t+1}],$$

$$\beta_{T_{вих},0} = -9.6 \cdot 10^{-7},$$



$$\beta_{T_{вих}} = [0.656 \quad 0.199 \quad 0.170 \quad 0.244 \quad 0.106],$$

$$\beta_{T_{газ,0}} = -6.7 \cdot 10^{-6},$$

$$\beta_{T_{газ}} = [0.302 \quad 0.691 \quad -0.195 \quad 2.09 \quad 1.39]$$

Окрім значень і змін вхідних параметрів на поточному кроці, також важливими можуть бути їх абсолютні значення, а також зміни, на минулих кроках. Якщо взяти їх до уваги, отримуємо залежність (9) та (10).

$$\Delta T_{вих, t+1} = \beta_{T_{вих,0}} + \sum_{i=1}^{p-1} \beta_{T_{вих,i+2}} \cdot \Delta Q_{t-i} + \sum_{j=1}^2 \sum_{i=0}^{p-1} \delta_{T_{вих,j,i+1}} \cdot x_{in,j,t-i}, \# \quad (9)$$

$$\Delta T_{газ, t+1} = \beta_{T_{вих,0}} + \sum_{i=1}^{p-1} \beta_{T_{вих,i+2}} \cdot \Delta Q_{t-i} + \sum_{j=1}^2 \sum_{i=0}^{p-1} \delta_{T_{вих,j,i+1}} \cdot x_{in,j,t-i}, \quad (10)$$

$$x_{in} = [\Delta T_{вих}, \Delta T_{вх}, \Delta T_{газ}]$$

де  $p$  – кількість попередніх кроків

$\beta_i, \delta_{j,i}$  – матриці коефіцієнтів вільних параметрів

Для визначення набору попередніх кроків  $p$ , було розраховано коефіцієнти лінійної регресії з  $p = 0 \dots 10$ . Було оцінено точність розрахунку температур при кожному з наборів вхідних параметрів за допомогою середньої квадратичної похибки.

Як видно з рис. 4 і рис. 5, середня квадратична похибка розрахунку обох параметрів спадає на всьому вибраному діапазоні значень  $p$ , але після  $p = 6$  цей спад є практично непомітним. Отримані коефіцієнти регресії:

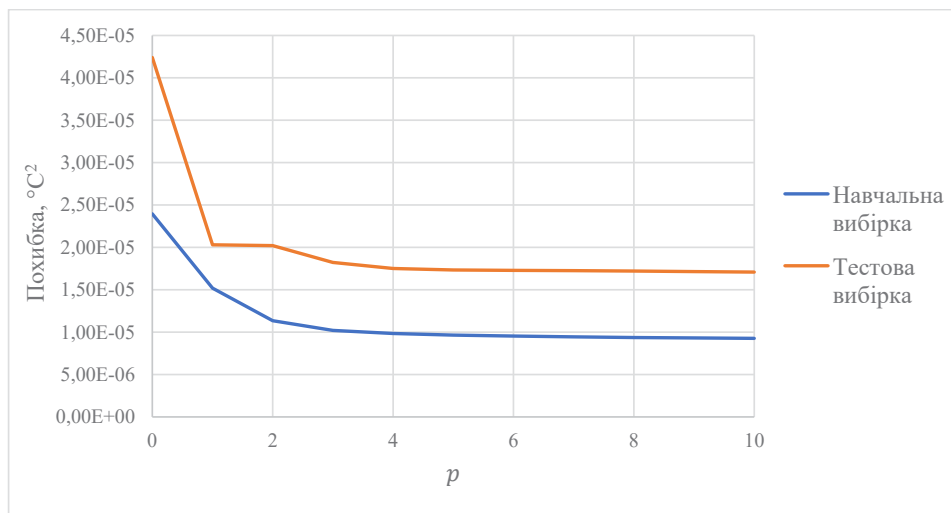


Рис. 4. Середня квадратична похибка розрахунку  $\Delta T_{вих,t+1}$  лінійної регресії в залежності від кількості попередніх кроків  $p$

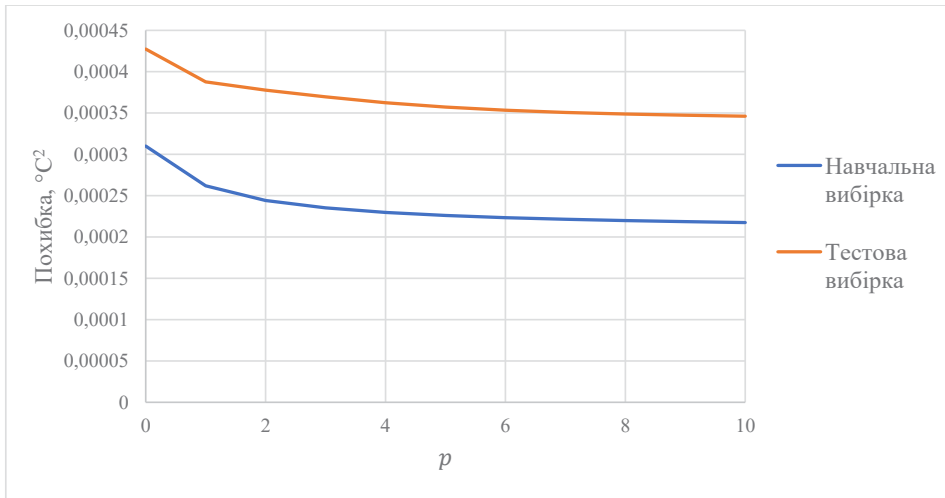


Рис. 5. Середня квадратична похибка розрахунку  $\Delta T_{газ,t+1}$  лінійної регресії в залежності від кількості попередніх кроків  $p$

$$\beta_{T_{вих},0} = 1.62 \cdot 10^{-6},$$

$$\beta_{T_{вих}} = \begin{bmatrix} 0.106 & 0.335 & 0.375 & 0.301 \\ 0.234 & 0.183 & 0.129 & 0.0674 \end{bmatrix},$$

$$\delta_{T_{вих}} = \begin{bmatrix} -0.286 & 1.99 & 0.178 \\ -0.0186 & -1.88 & 0.157 \\ 0.0821 & 0.0912 & 0.132 \\ 0.101 & -1.61 & 0.0553 \\ 0.0871 & 3.92 & 0.0144 \\ 0.0654 & -1.76 & -0.00266 \\ 0.0306 & -0.284 & -0.00174 \end{bmatrix},$$

$$\beta_{T_{газ},0} = -7.87 \cdot 10^{-6},$$

$$\beta_{T_{газ}} = \begin{bmatrix} 1.39 & 2.73 & 2.18 & 1.72 \\ 1.32 & 0.948 & 0.604 & 0.290 \end{bmatrix}$$

$$\delta_{T_{газ}} = \begin{bmatrix} -0.524 & 2.57 & -0.591 \\ -0.487 & -2.07 & -0.269 \\ -0.287 & 3.04 & -0.0648 \\ -0.0542 & -14.1 & 0.0379 \\ 0.158 & 25.2 & 0.0535 \\ 0.253 & -7.83 & 0.0315 \\ 0.167 & -4.58 & 0.0166 \end{bmatrix},$$

Порівняємо похибку розрахунку з похибкою лінійної регресії до ранжування вхідних параметрів з похибкою до введення попередніх кроків.

Таблиця 3

**Порівняння похибок прогнозування лінійною регресією**

Алгоритм	Середня квадратична похибка при прогнозуванні			
	$\Delta T_{вих,t+1}$		$\Delta T_{газ,t+1}$	
	навчальна вибірка	тестова вибірка	навчальна вибірка	тестова вибірка
Лінійна регресія	2.35e-5	4.21e-5	3.07e-4	4.29e-4
Лінійна регресія після ранжування входів	2.40e-5	4.24e-5	3.10e-4	4.27e-4
Лінійна регресія після ранжування входів з введенням попередніх кроків	9.54e-6	1.73e-5	2.23e-4	3.53e-4

З результатів у табл. 3, після ранжування вхідних параметрів, похибка розрахунку збільшилась, в середньому, на 0.8%, при тому що кількість параметрів, а отже і складність розрахунку, зменшилась з 10 до 5.

Після введення параметрів на попередніх кроках, у порівнянні з початковою лінійною регресією, похибка розрахунку  $\Delta T_{вих,t+1}$  зменшилась на 59%, а похибка розрахунку  $\Delta T_{газ,t+1}$  зменшилась на 23%. Це є значним покращенням, але при цьому кількість вхідних параметрів зросла до 29.

#### 5. Регресивне дерево рішень

Введення нелінійних залежностей у модель лінійної регресії (наприклад, залежностей від взаємодії параметрів) дуже сильно збільшує складність розрахунку такого рівняння: поліном другого степеню від 10 параметрів буде нараховувати 66 параметрів, а третього – 258. Причиною цьому є те, що лінійна регресія є глобальною моделлю, яка покриває всі випадки однією формулою.

Іншим підходом до вирішення нелінійної задачі є розбиття простору даних на менші частини, в яких задача буде легшою [16]. Після цього, ці частини можна розбити ще – це називається рекурсивним розбиттям.

Одним із способів вирішення задачі рекурсивним розбиттям є побудова регресивного дерева [17]. Дерево складається з вузлів розбиття і кінцевих вузлів (листя). Кожен вузол позначений питанням, відповідь на яке вирішує по якій з наступних гілок піде рішення. При чому питання є логічним виразом (має відповідь або «так», або «ні») і стосується лише одного параметру. Листя є підмоделями рекурсивного розбиття, які у регресивних деревах є постійними значеннями.

Побудова дерева складається з рекурсивного повторення розбиття. При розбитті, вибирається питання, яке максимізує кількість інформації, яка отримується при цьому. При чому, аби не сталось ситуації, коли кожна комбінація вхідних і вихідних параметрів лежить в окремому кінцевому вузлі, вводиться порогове значення  $\delta$ . У випадку якщо інформаційне навантаження розбиття менше ніж  $\delta$ , розбиття не буде – цей вузол стане листком. Як інформаційність дерева береться сума квадратів помилок (6). Розбиття має зменшувати цю суму на достатню величину.

$$S = \sum_{c \in \text{leaves}(T)} \sum_{i \in C} (y_i - m_c)^2, \quad (11)$$

$$m_c = \frac{1}{n_c} \sum_{i \in C} y_i$$

де  $leaves(T)$  – множина всіх кінцевих вузлів дерева  $T$

$m_c$  – прогнозоване значення кінцевого вузла

$C$  – множина точок даних у поточному вузлі

$y_i$  – прогнозована змінна

Отже рекурсивне розбиття складається з повторення таких дій, починаючи зі створення першого вузла дерева:

1. Обчислити  $m_c$  та  $S$  для дерева.

2. Якщо у всіх точках даних вузла всі незалежні змінні однакові, зупинитись. У іншому випадку, провести пошук по всім можливим бінарним розбиттям усіх змінних для знаходження такого, яке зменшить  $S$  більше всього. Якщо зменшення  $S$  буде менше порогу  $\delta$ , зупинитись. В іншому випадку, зробити розбиття.

3. Для кожного з нових вузлів, перейти в пункт 1.

Прогнозування таким деревом відбувається просто слідуванням розбиттям залежно від вхідних змінних.

Було побудовано групу регресійних дерев для набору даних з попередніми кроками. Древа не обмежувалось зі сторони мінімального приросту точності  $\delta$ , але аби контролювати розмір дерев, було обмежено їх глибину. Оцінимо точність прогнозування деревами з глибинами від 1 до 40 за допомогою середньої квадратичної та середньої абсолютної похибок.

Як видно з рис. 6, середня квадратична і середня абсолютна похибки розрахунку  $\Delta T_{вих,t+1}$  спадають до глибини дерева 25, після чого трохи збільшуються. Оберемо це число як достатню глибину для розрахунку даного параметру. З рис. 7 видно, що середня квадратична похибка розрахунку  $\Delta T_{газ,t+1}$  спадає до глибини дерева 18, але середня абсолютна похибка перестає спадати лише при глибині 30. Оберемо це число як достатню глибину для розрахунку даного параметру. Порівняємо похибку розрахунку з похибкою лінійної регресії після ранжування вхідних параметрів і введення попередніх кроків.

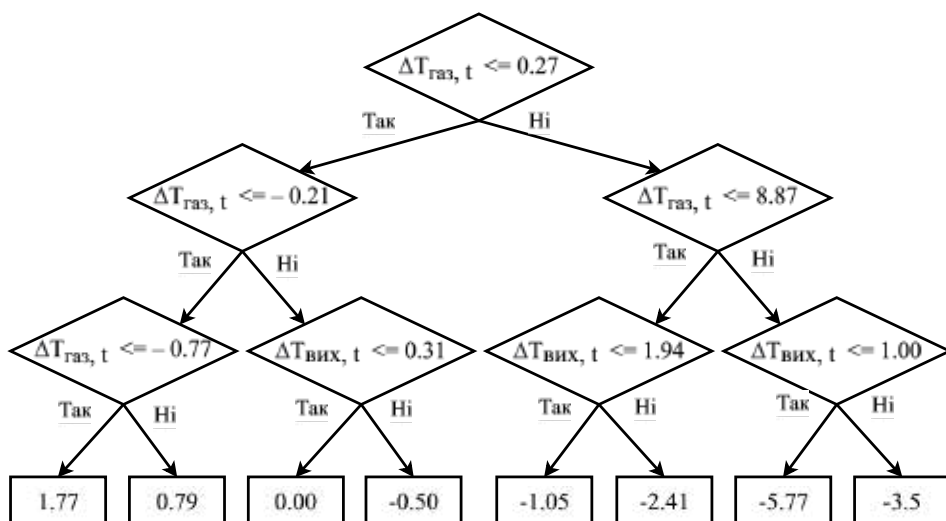


Рис. 6. Регресивне дерево розрахунку  $\Delta T_{вих,t+1}$  з глибиною 4

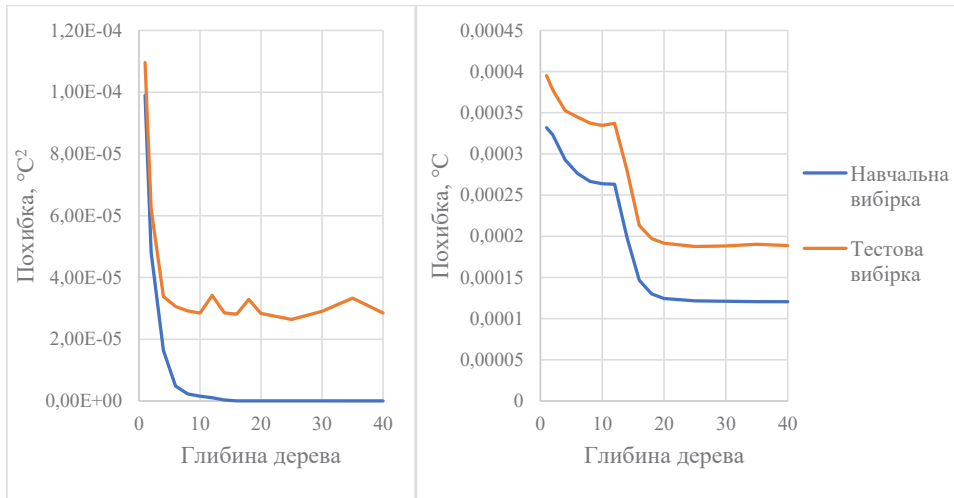


Рис. 7. Середня квадратична та середня абсолютна похибки розрахунку  $\Delta T_{вих,t+1}$  регресійного дерева в залежності від глибини дерева

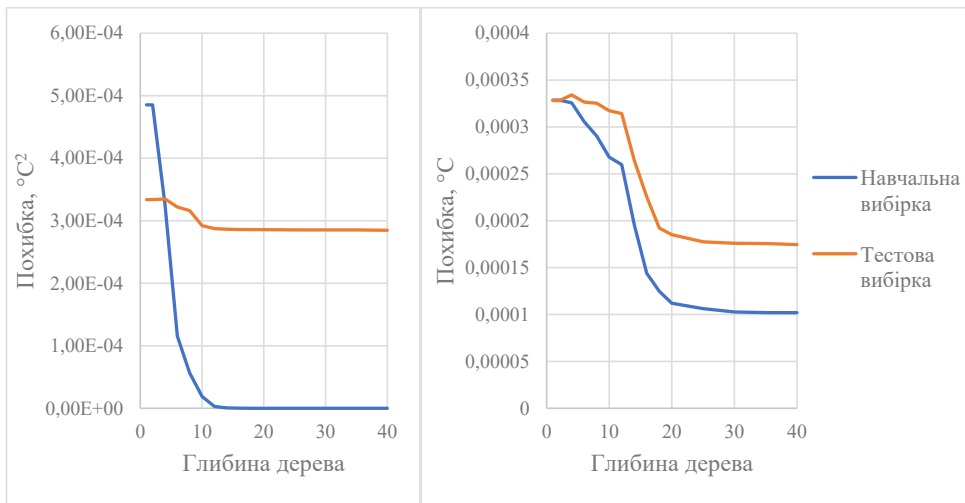


Рис. 8. Середня квадратична похибка розрахунку  $\Delta T_{газ,t+1}$  регресійного дерева в залежності від глибини дерева

Як видно з результатів у табл. 4, у розрахунку  $\Delta T_{вих,t+1}$ , у порівнянні з лінійною регресією, регресивне дерево на декілька порядків зменшило похибку розрахунку на навчальній вибірці, але на тестовій вибірці похибка збільшилась на 52%. У розрахунку  $\Delta T_{газ,t+1}$ , як і у минулому випадку, похибка на навчальній вибірці зменшилась значно, але на тестовій вибірці в похибка зменшилась на 24%.

Таблиця 4

**Порівняння похибок прогнозування регресивним деревом  
і лінійною регресією**

Алгоритм	Середня квадратична похибка при прогнозуванні			
	$\Delta T_{вих,t+1}$		$\Delta T_{газ,t+1}$	
	навчальна вибірка	тестова вибірка	навчальна вибірка	тестова вибірка
Регресивне дерево	4.03e-8	2.64e-5	3.07e-8	2.85e-4
Лінійна регресія після ранжування входів з введенням попередніх кроків	9.54e-6	1.73e-5	2.23e-4	3.53e-4

**6. Ансамблевий метод «Голосування»**

Різні типи регресивних моделей, або навіть однакові типи моделей, розраховані на різних частинах вхідних даних, значно краще виокремлюють деякі, унікальні у кожному випадку, залежності між вхідними і вихідними змінними. Поєднання декількох екземплярів регресивних моделей одного або різних типів називається ансамблевими методами машинного навчання.

Найпростішим ансамблевим методом є «голосування» [18]. Суть цього методу полягає в усередненні прогнозів декількох моделей для покращення точності. Прогнози раніше розрахованих лінійної регресії і регресивного дерева було усереднено та оцінено точність такої моделі за допомогою середньої квадратичної похибки та порівняємо з похибками окремих моделей.

Таблиця 5

**Порівняння похибок прогнозування методом «голосування»  
з регресивним деревом і лінійною регресією**

Алгоритм	Середня квадратична похибка при прогнозуванні			
	$\Delta T_{вих,t+1}$		$\Delta T_{газ,t+1}$	
	навчальна вибірка	тестова вибірка	навчальна вибірка	тестова вибірка
Метод «голосування»	2.40e-6	2.03e-5	5.59e-5	2.31e-4
Регресивне дерево	4.03e-8	2.64e-5	3.07e-8	2.85e-4
Лінійна регресія після ранжування входів з введенням попередніх кроків	9.54e-6	1.73e-5	2.23e-4	3.53e-4

Як видно з результатів у табл. 5, у розрахунку  $\Delta T_{вих,t+1}$ , метод «голосування» усереднив похибки розрахунків окремих алгоритмів: на навчальній вибірці похибка зменшилась у 4 рази відносно результатів лінійної регресії, а на тестовій вибірці похибка зменшилась на 30% відносно результатів регресивного дерева. У розрахунку  $\Delta T_{газ,t+1}$ , відбулось майже те-саме: на навчальній вибірці похибка зменшилась у 4 рази відносно результатів лінійної регресії, а на тестовій вибірці похибка зменшилась як відносно результатів лінійної регресії, так і відносно результатів регресивного дерева: на 53% і 23% відповідно.

Модель було застосовано для багатокрокового моделювання на основі частки даних з валідаційної вибірки. Фінальні похибки такого моделювання: середня квадратична –  $0.077\text{ }^{\circ}\text{C}^2$ , середня абсолютна –  $0.13\text{ }^{\circ}\text{C}$ .

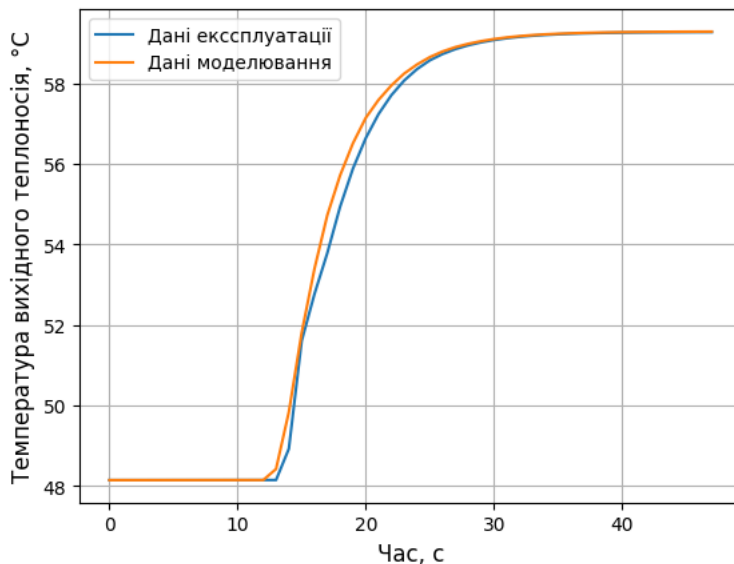


Рис. 9. Порівняння перехідних процесів з реальних даних та моделювання

**Висновки.** Розроблено регресійну модель температур вихідної води і димових газів конденсаційного котла. Проаналізовано регресивні моделі машинного навчання для вирішення задачі розрахунку температур вихідного теплоносія і димових газів. На основі аналізу було створено модель лінійної регресії, після чого було зменшено набір вхідних параметрів і параметрів на минулих кроках для досягнення компромісу між точністю моделі і складністю її розрахунку. Після того, було розраховано модель регресивного дерева і методу «голосування» – поєднання двох попередніх моделей. Похибка останньої повністю задовольняє вимогам точності. Надалі, розрахована статистична модель буде використана у цифровому двійнику.

У порівнянні з класичними методами ідентифікації та моделювання об'єктів на основі диференціальних рівнянь або передатних функцій, використання методів машинного навчання забезпечує швидше і зручніше отримання моделі завдяки можливості автоматичного оброблення великого обсягу даних та виявлення прихованих закономірностей без необхідності в ручному налаштуванні моделі, забезпечуючи високу точність при мінімальних обчислювальних витратах.

#### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ:

1. Національна економічна стратегія 2030 / Платформа Центру економічного відновлення. URL: <https://nes2030.org.ua/> (Дата звернення: 30.05.2024).
2. Патон Б.Є., Долінський А.А., Геєць В.М., Кухар В.П., Басок Б.І., Базєєв Є.Т., Подолець Р.З. Пріоритети Національної стратегії теплозабезпечення населених пунктів України. *Вісник НАН України*. 2014. № 9. С. 29-47.

3. Harish Satyavada, Simone Baldi Monitoring energy efficiency of condensing boilers via hybrid first-principle modelling and estimation. *Energy*, vol. 142, 2018, pp. 121-129. URL: <https://doi.org/10.1016/j.energy.2017.09.124> (Дата звернення: 30.05.2024).
4. João Barros, Condenser boiler modeling. URL: <https://fenix.tecnico.ulisboa.pt/downloadFile/395144950949/resumo.pdf> (Дата звернення: 30.05.2024)
5. Куделя П. П., Барабаш П. А. Тепловой баланс конденсационных котлов. *Энергетика*. 2014. № 1. С. 51-59. URL: [http://nbuv.gov.ua/UJRN/eete\\_2014\\_1\\_9](http://nbuv.gov.ua/UJRN/eete_2014_1_9) (Дата звернення: 30.05.2024)
6. Malakuti Somayeh, Van Schalkwyk Pieter, Boss Birgit, Sastry Chellury, Runkana Venkataramana, Lin Shi-Wan, Rix Simon, Green Gavin, Baechle Kilian, Nath Shyam. Digital Twins for Industrial Applications. Definition, Business Values, Design Aspects, Standards and Use Cases. *An Industrial Internet Consortium White Paper. Version 1.0*. 2020. URL: [https://www.iiconsortium.org/pdf/IIC\\_Digital\\_Twins\\_Industrial\\_Apps\\_White\\_Paper\\_2020-02-18.pdf](https://www.iiconsortium.org/pdf/IIC_Digital_Twins_Industrial_Apps_White_Paper_2020-02-18.pdf) (Дата звернення: 30.05.2024).
7. Harish Satyavada, Simone Baldi. A Novel Modelling Approach for Condensing Boilers Based on Hybrid Dynamical Systems. Delft Center for Systems and Control. *Delft University of Technology*. 2016. vol. 4(2), no. 10. URL: <https://doi.org/10.3390/machines4020010> (Дата звернення: 22.04.2024).
8. Dan Teodor Bălănescu, Vlad Mario. Homutescu Experimental investigation on performance of a condensing boiler and economic evaluation in real operating conditions. *Applied Thermal Engineering*. 2018. Volume 143. p. 48-58. URL: <https://doi.org/10.1016/j.applthermaleng.2018.07.082> (Дата звернення: 30.05.2024).
9. Hanby, V. I. Modelling the performance of condensing boilers. *Journal of the Energy Inst.* 2007. № 80. p. 229-231.
10. Дубовой В. М. Ідентифікація та моделювання технологічних об'єктів і систем керування : навчальний посібник. Вінниця : ВНТУ, 2012. 308 с.
11. Кубрак А. І., Жученко А. І., Кваско М. З. Комп'ютерне моделювання та ідентифікація автоматичних систем. Київ: Політехніка, 2004. 424 с.
12. Жученко А. І. Ідентифікація динамічних характеристик. Комп'ютерні методи. Київ, 2000. 182 с. URL: [http://opac.kpi.ua/F/?func=direct&doc\\_number=000066839&local\\_base=КРІ01](http://opac.kpi.ua/F/?func=direct&doc_number=000066839&local_base=КРІ01) (Дата звернення: 22.04.2024).
13. Юзвак С.В., Шемчук В.В., Самарін С.В., Косулько А.Г., Лахно В.Г. Модульні котельні установки системи «УКРІНТЕРМ». Довідковий посібник (12-е вид.). УКРІНТЕРМ. Біла Церква, 2019. 171 с.
14. Machine learning regression algorithms to predict emissions from steam boilers / В. D. Ross-Veitía та ін. *Heliyon*. 2024. e26892. URL: <https://doi.org/10.1016/j.heliyon.2024.e26892> (дата звернення: 02.06.2024).
15. Afyouni S., Smith S. M., Nichols T. E. Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation. *NeuroImage*. 2019. vol. 199. pp. 609–625. URL: <https://doi.org/10.1016/j.neuroimage.2019.05.011> (дата звернення: 02.06.2024).
16. Chou P. A. Optimal partitioning for classification and regression trees. *IEEE transactions on pattern analysis and machine intelligence*. 1991. Vol. 13, no. 4. pp. 340–354. URL: <https://doi.org/10.1109/34.88569> (дата звернення: 02.06.2024).
17. Carnegie Mellon University, Department of Statistics & Data Science. "Statistics 36-350: Data Mining, Fall 2006." Lecture notes. 2006. Lecture 10 – Regressive trees. URL: <https://www.stat.cmu.edu/~cshalizi/350-2006/lecture-10.pdf> дата звернення: 30.05.2024).
18. Liapis C. M., Karanikola A., Kotsiantis S. Data-Efficient software defect prediction: a comparative analysis of active learning-enhanced models and voting ensembles. *Information sciences*. 2024. no. 120786. URL: <https://doi.org/10.1016/j.ins.2024.120786> (дата звернення: 02.06.2024).



## REFERENCES:

1. Platforma Tsentru ekonomichnoho vidnovlennya (n.d.). *Natsionalna ekonomichna stratehiya 2030* [National Economic Strategy 2030]. Retrieved from: <https://nes2030.org.ua/> (accessed 30 May 2024).
  2. Paton B.E., Dolinskyi A.A., Heyets V.M., Kukhar V.P., Basok B.I., Bazyeyev Y.T., Podolets R.Z. (2014) Prioritytety Natsionalnoyi stratehiyi teplozabezpechennya naselenykh punktiv Ukrayiny [Target reference points of the national strategy for heating supplies of settlements of Ukraine]. *Visnyk NAN Ukrayiny*, no. 9. pp. 29-47.
  3. Harish Satyavada, Simone Baldi (2018) Monitoring energy efficiency of condensing boilers via hybrid first-principle modelling and estimation. *Energy*, vol. 142, pp. 121-129. <https://doi.org/10.1016/j.energy.2017.09.124>
  4. João Barros (n.d.). *Condenser boiler modeling*. Retrieved from: <https://fenix.tecnico.ulisboa.pt/downloadFile/395144950949/resumo.pdf> (accessed 30 May 2024).
  5. Kudelya P. P., Barabash P. A. (2014) Teplovoy balans kondensatsionnykh kotlov [Thermal balance for condensing boilers]. *Enerhetyka*, no. 1. pp. 51-59. Retrieved from: [http://nbuv.gov.ua/UJRN/eete\\_2014\\_1\\_9](http://nbuv.gov.ua/UJRN/eete_2014_1_9) (accessed 30 May 2024)
  6. Malakuti Somayeh, Van Schalkwyk Pieter, Boss Birgit, Sastry Chellury, Runkana Venkataramana, Lin Shi-Wan, Rix Simon, Green Gavin, Baechle Kilian, Nath Shyam (2020) Digital Twins for Industrial Applications. Definition, Business Values, Design Aspects, Standards and Use Cases. *An Industrial Internet Consortium White Paper. Version 1.0*. Retrieved from: [https://www.iiconsortium.org/pdf/IIC\\_Digital\\_Twins\\_Industrial\\_Apps\\_White\\_Paper\\_2020-02-18.pdf](https://www.iiconsortium.org/pdf/IIC_Digital_Twins_Industrial_Apps_White_Paper_2020-02-18.pdf) (accessed 30 May 2024).
  7. Harish Satyavada, Simone Baldi (2016) A Novel Modelling Approach for Condensing Boilers Based on Hybrid Dynamical Systems. Delft Center for Systems and Control. *Delft University of Technology*, vol. 4(2), no. 10. <https://doi.org/10.3390/machines4020010>
  8. Dan Teodor Bălănescu, Vlad Mario. Homutescu (2018) Experimental investigation on performance of a condensing boiler and economic evaluation in real operating conditions. *Applied Thermal Engineering*. vol. 143. pp. 48-58. <https://doi.org/10.1016/j.applthermaleng.2018.07.082>
  9. Hanby V. I. (2007) Modelling the performance of condensing boilers. *Journal of the Energy Inst*, no. 80, pp. 229-231.
  10. Dubovoy V. M. (2012) *Identyfikatsiya ta modelyuvannya tekhnolohichnykh obyektiv i system keruvannya : navchalnyy posibnyk* [Identification and modeling of technological objects and control systems: a study guide]. Vinnytsya: VNTU. (in Ukrainian)
  11. Kubrak A.I., Zhuchenko A. I., Kvasko M. Z. (2004) *Kompyuterne modelyuvannya ta identyfikatsiya avtomatichnykh system* [Computer modeling and identification of automatic systems]. Kyiv: Politekhnika. (in Ukrainian)
  12. Zhuchenko A. I., Kvasko M. Z., Kubrak A. I. (2000) *Identyfikatsiya dynamichnykh kharakterystyk. Kompyuterni metody* [Identification of dynamic characteristics. Computer methods]. Kyiv: VIPOL. (in Ukrainian)
  13. Yuzvak S.V., Shemchuk V.V., Samarin S.V., Kosulko A.H., Lakhno V.H. (2019) *Modulni kotelni ustanovky systemy «UKRINTERM»*. *Dovidkovyy posibnyk (12-e vyd.)* [Modular boiler installations of the "UKRINTERM" system. Reference Manual (12th ed.)]. Bila Tserkva: UKRINTERM. (in Ukrainian)
  14. B. D. Ross-Veitia, R. Arias-Gilart, R. E. Conde-García, A. Espinel-Hernández, J. R. Nuñez-Alvarez, H. Hernández-Herrera, Y. E. Llosas-Albuerno (2024) Machine learning regression algorithms to predict emissions from steam boilers. *Heliyon*, vol. 10, issue5, e26892. <https://doi.org/10.1016/j.heliyon.2024.e26892>
  15. Afyouni S., Smith S. M., Nichols T. E. (2019) Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation. *NeuroImage*, vol. 199. pp. 609–625. <https://doi.org/10.1016/j.neuroimage.2019.05.011>
-

16. Chou P. A. (1991) Optimal partitioning for classification and regression trees. *IEEE transactions on pattern analysis and machine intelligence*, vol. 13, no. 4. pp. 340–354. <https://doi.org/10.1109/34.88569>

17. Carnegie Mellon University, Department of Statistics & Data Science (2006) Lecture 10 – Regressive trees. 36-350: Data Mining. Retrieved from: <https://www.stat.cmu.edu/~cshalizi/350-2006/lecture-10.pdf> (accessed 30 May 2024).

18. Liapis C. M., Karanikola A., Kotsiantis S. (2024) Data-Efficient software defect prediction: a comparative analysis of active learning-enhanced models and voting ensembles. *Information sciences*, no. 120786. <https://doi.org/10.1016/j.ins.2024.120786>.