

УДК 004.77

DOI <https://doi.org/10.32689/maup.it.2025.2.21>

Олексій ПІСКАРЬОВ

кандидат технічних наук, доцент,

доцент кафедри електронних обчислювальних машин,

Харківський національний технічний університет радіоелектроніки, oleksii.piskarov@nure.ua

ORCID: 0000-0002-6980-984X

Данило АБРАМОВИЧ

здобувач РВО магістр-науковець,

Харківський національний технічний університет радіоелектроніки, danylo.abramovych1@nure.ua

ORCID: 0009-0007-0500-4683

Владислав ПЛУГІН

здобувач РВО магістр-науковець,

Харківський національний технічний університет радіоелектроніки, vladyslav.pluhin@nure.ua

ORCID: 0009-0002-1242-6378

ЗАСТОСУВАННЯ НЕЙРОМЕРЕЖ ТА ГІБРИДНИХ СХОВИЩ ПРИ СТВОРЕННІ КОМП'ЮТЕРНИХ СИСТЕМ

Анотація. У сучасних комп'ютерних системах зростає потреба у швидкій обробці, надійному зберіганні та точному аналізі великих обсягів даних у реальному часі. Це особливо актуально для автоматизованих технологічних процесів, інтелектуальних середовищ та керування розподіленими об'єктами. У таких умовах традиційні архітектури зберігання та аналітики даних стають недостатньо гнучкими та масштабованими.

Метою статті є огляд сучасних підходів до побудови комп'ютерних систем прогнозування стану динамічних об'єктів шляхом поєднання методів машинного навчання з гібридною архітектурою зберігання інформації. У рамках роботи здійснюється оцінка точності різних моделей штучного інтелекту для короткострокового прогнозування параметрів, а також аналізується ефективність використання гібридних сховищ даних для забезпечення надійного, масштабованого та продуктивного функціонування таких систем у реальному часі.

Методи дослідження: використовуються методи машинного навчання, зокрема регресійні та ансамблеві моделі, а також експериментальне моделювання на основі часових рядів. Для оцінки ефективності застосовано метрики точності та аналіз впливу вхідних параметрів. Обробка й зберігання даних реалізовані з використанням гібридної інфраструктури.

Наукова новина дослідження полягає в тому, що проведений аналіз виявив суттєві перспективи інтеграції високоточних моделей машинного навчання з гібридними сховищами інформації для створення адаптивної комп'ютерної системи прогнозування, здатної до оперативної реакції та довготривалої аналітики в умовах обробки великих обсягів даних у реальному часі.

Висновки. Проведене дослідження довело доцільність використання машинного навчання у поєднанні з гібридною архітектурою зберігання даних для створення інтелектуальних комп'ютерних систем прогнозування. На основі експериментального моделювання встановлено, що модель XGBoost демонструє найвищу точність, стабільність результатів і здатність до інтерпретації впливу вхідних параметрів на вихідні дані, значно випереджаючи інші протестовані алгоритми. Важливою складовою запропонованого підходу стала побудова гібридного сховища інформації, яке поєднує переваги локального зберігання, NoSQL-систем та хмарної інфраструктури. Така структура дозволяє забезпечити як оперативне реагування на зміни вхідних параметрів, так і проведення глибокого історичного аналізу для довготривалого планування та вдосконалення алгоритмів. Запропонована система може бути адаптована для широкого кола завдань у сфері автоматизованого керування, цифрового моніторингу, управління виробничими процесами та в інших прикладних напрямках. Вона дозволяє не лише підвищити ефективність роботи комп'ютерних систем, а й закладає основу для створення самонавчальних, адаптивних рішень нового покоління, здатних працювати в умовах реального часу з великими обсягами даних.

Ключові слова: комп'ютерні системи, машинне навчання, гібридні сховища, інтелектуальне керування, автоматизація, хмарні технології.

Oleksiy PISKAROV, Danilo ABRAMOVICH, Vladislav PLUHIN. APPLICATION OF NEURAL NETWORKS AND HYBRID STORAGE IN THE CREATION OF COMPUTER SYSTEMS

Abstract. In modern computer systems, there is a growing need for fast processing, reliable storage, and accurate analysis of large amounts of data in real time. This is especially true for automated technological processes, intelligent environments, and distributed object management. In such circumstances, traditional data storage and analytics architectures become insufficiently flexible and scalable. The purpose of the article is to review modern approaches to building computer systems

© О. Піскарів, Д. Абрамович, В. Плугін, 2025

Стаття поширюється на умовах ліцензії CC BY 4.0

for predicting the state of dynamic objects by combining machine learning methods with a hybrid information storage architecture. The paper evaluates the accuracy of various artificial intelligence models for short-term parameter forecasting and analyzes the effectiveness of using hybrid data warehouses to ensure reliable, scalable, and productive operation of such systems in real time.

Research methods: The study uses machine learning methods, including regression and ensemble models, as well as experimental modeling based on time series. Accuracy metrics and analysis of the influence of input parameters are used to evaluate the efficiency. Data processing and storage are implemented using a hybrid infrastructure.

The scientific novelty of the study is that the analysis revealed significant prospects for integrating high-precision machine learning models with hybrid information warehouses to create an adaptive computer forecasting system capable of rapid response and long-term analytics in the context of processing large amounts of data in real time.

Conclusions. The study proved the feasibility of using machine learning in combination with a hybrid data storage architecture to create intelligent computer forecasting systems. Based on experimental modeling, it was found that the XGBoost model demonstrates the highest accuracy, stability of results, and ability to interpret the impact of input parameters on output data, significantly outperforming other tested algorithms. An important component of the proposed approach was the construction of a hybrid information storage that combines the advantages of local storage, NoSQL systems, and cloud infrastructure. Such a structure allows for both rapid response to changes in input parameters and in-depth historical analysis for long-term planning and algorithm improvement. The proposed system can be adapted for a wide range of tasks in the field of automated control, digital monitoring, production process control, and other applications. It allows not only to increase the efficiency of computer systems, but also lays the foundation for the creation of self-learning, adaptive solutions of a new generation capable of working in real time with large amounts of data.

Key words: computer systems, machine learning, XGBoost, forecasting, hybrid storage, real-time data processing, intelligent control, automation, NoSQL, cloud technologies.

Постановка проблеми та аналіз останніх досліджень і публікацій. Сучасні комп'ютерні системи функціонують в умовах зростаючої складності керованих об'єктів та стрімкого збільшення обсягів даних, що потребують обробки в реальному часі. Ефективне функціонування таких систем передбачає не лише оперативну реакцію на зміну вхідних параметрів, але й здатність до адаптивного прогнозування подальших станів об'єкта. Усе це вимагає впровадження нових архітектур обробки й збереження інформації та інтелектуальних методів аналізу даних. Окремим викликом є потреба в одночасному забезпеченні швидкодії, масштабованості, точності та надійності систем зберігання та обробки інформації. Традиційні централізовані рішення або класичні реляційні бази даних часто не відповідають вимогам до гнучкості й продуктивності, особливо в умовах динамічної зміни даних та високої інтенсивності запитів [2]. У зв'язку з цим актуальним є використання гібридних сховищ інформації, які поєднують локальні обчислення, розподілені бази даних й хмарну інфраструктуру. Поєднання таких сховищ із методами машинного навчання створює основу для побудови комп'ютерних систем, здатних не тільки накопичувати та обробляти інформацію, а й здійснювати точне прогнозування на основі історичних і поточних даних.

Постає завдання розробити підхід, який дозволяє інтегрувати інтелектуальні моделі прогнозування з адаптивною системою зберігання даних, орієнтованою на роботу в умовах великої кількості змінних параметрів та високої частоти оновлення інформації. Тема інтеграції інтелектуальних моделей прогнозування з інфраструктурою зберігання даних активно досліджується в останні роки у контексті цифрової трансформації, автоматизації керування та розвитку систем реального часу. Значна кількість праць зосереджена на застосуванні алгоритмів машинного навчання, зокрема моделей регресії, опорних векторів, нейронних мереж і градієнтного бустингу, для вирішення задач прогнозування у динамічних інформаційних середовищах. У працях, присвячених хмарним й edge-системам, доведено переваги комбінованого зберігання даних, де частина інформації обробляється локально, а інша передається у хмару для аналітики або навчання моделей [14]. Такий підхід дозволяє скоротити час реакції системи та мінімізувати навантаження на центральні сервери.

Попри значний прогрес у дослідженні окремих компонентів – моделей машинного навчання або архітектур зберігання даних – комплексна інтеграція прогнозних алгоритмів у гібридну інфраструктуру все ще залишається відкритим і перспективним напрямом. Таким чином існує потреба в дослідженнях, що розглядатимуть цю інтеграцію як єдину систему з урахуванням вимог до реального часу, масштабованості та точності прогнозування.

Метою цієї роботи є огляд методів та підходів до розробки й експериментальної перевірки архітектури комп'ютерної системи прогнозування динамічних параметрів, яка поєднує сучасні алгоритми машинного навчання з гібридною інфраструктурою зберігання даних. Такий підхід забезпечує обробку великих обсягів інформації у реальному часі, підвищує точність прогнозів, знижує навантаження на обчислювальні ресурси та гарантує гнучкість роботи в умовах змін вхідних параметрів. Особлива увага приділена побудові універсальної та масштабованої структури, здатної до збереження контексту даних і автономної роботи з високою надійністю.

Сучасні вимоги до комп'ютерних систем управління зумовлюють відхід від централізованих моделей на користь адаптивних та інтелектуальних рішень. Інтеграція штучного інтелекту, особливо машинного навчання, з гібридними сховищами дозволяє не лише накопичувати дані, а й здійснювати їх аналітичну обробку у реальному часі з подальшим адаптивним реагуванням [6].

На відміну від традиційних фізичних і статистичних моделей, алгоритми boosting (зокрема XGBoost) демонструють високу точність прогнозування завдяки здатності виявляти складні нелінійні взаємозв'язки. Крім ефективності, ці моделі зручні для інтерпретації, що дозволяє аналізувати значущість вхідних параметрів за допомогою вагових коефіцієнтів та F-скорів [11].

XGBoost вирізняється швидким навчанням, підтримкою паралельних обчислень і зменшеними вимогами до ресурсів, що робить його придатним для систем із жорсткими обмеженнями на обчислювальні потужності. На відміну від глибоких нейронних мереж, що потребують потужного обладнання (GPU), boosting-моделі можуть працювати на звичайних машинах без втрати точності [11]. Це робить XGBoost актуальним для побудови адаптивних комп'ютерних систем управління з пояснюваними прогнозами.

Одним із прикладів застосування такого підходу є комп'ютерна система регулювання мікроклімату у великих приміщеннях, де критичне значення має точне прогнозування параметрів середовища. Такі об'єкти, як аграрні сховища чи виробничі комплекси, оснащуються системами моніторингу й управління – від сенсорів до виконавчих механізмів – для підтримки стабільних умов [4]. Однак класичні підходи (на кшталт FDM, FEM) мають обмеження в адаптації до змін і високої варіативності параметрів.

Інтеграція нейронних мереж із гібридними сховищами відкриває нові можливості для автоматизованого управління середовищем. З огляду на складність і вартість використання глибоких нейромереж, дедалі більше досліджень фокусуються на boosting-моделях, як-от XGBoost, які забезпечують високу точність та інтерпретованість, а також легко інтегруються з розподіленими платформами.

У сільському господарстві XGBoost ефективно використовується для прогнозування параметрів мікроклімату, зокрема температури, вологості та рівня CO₂, що безпосередньо впливають на врожайність і якість продукції. Поєднання цієї моделі з гібридними сховищами і сенсорними мережами дозволяє формувати адаптивні системи з точним прогнозуванням та швидким реагуванням у реальному часі.

Гібридне сховище даних сприяє самонавчанню моделей, завдяки механізмам оновлення параметрів за новими даними. Це підвищує адаптивність і точність системи, забезпечуючи надійність і стійкість до відмов, оскільки критичні дані дублюються в різних середовищах. Розподіленість і багаторівнева обробка інформації – від периферійних пристроїв до хмарних платформ – дозволяють зменшити затримки та навантаження на центральні обчислення.

Таким чином, впровадження гібридної архітектури із застосуванням XGBoost та нейронних мереж дозволяє створити ефективну систему автоматизованого керування мікрокліматом. Вона поєднує точність прогнозів, адаптивність, високу продуктивність і масштабованість – критично важливі для великих приміщень і складних технологічних процесів.

Виклад основного матеріалу. У цьому дослідженні застосування нейронних мереж базувалося на даних, зібраних у великому приміщенні площею 1320 м², яке оснащено сучасними системами моніторингу та автоматизованого управління параметрами внутрішнього середовища. Для вимірювання мікрокліматичних показників використовувалися сенсори, що фіксували температуру, відносну вологість, концентрацію CO₂, швидкість і напрям повітряних потоків, рівень сонячної радіації та опадів. Отримані значення слугували основою для побудови моделей машинного навчання, які оптимізували умови середовища відповідно до виробничих вимог [7].

Протягом 56 днів дані реєструвалися щохвилини й зберігалися у хмарній базі, що гарантувало високу точність фіксації змін. Внутрішні параметри середовища контролювалися сенсорами температури, вологості та газового складу (SH-300-DC), розміщеними у ключових зонах приміщення. Дані про активність обігріву надходили з контролера керування опаленням, що дозволяло детально аналізувати зв'язок між кліматичними умовами та роботою регуляційних механізмів. Інформація накопичувалася у гібридному сховищі, яке забезпечувало ефективну обробку великих масивів і швидку реакцію системи на зміну параметрів.

Зібрані характеристики середовища використовувалися як вхідні змінні для побудови прогнозних моделей. Для машинного навчання застосовувалися бібліотеки Python 3.8: Scikit-learn – для реалізації множинної лінійної регресії (MLR) та методу опорних векторів (SVR), а також TensorFlow і XGBoost – для побудови штучних нейронних мереж (ANN) і boosting-моделі відповідно.

Гіперпараметри моделей налаштовувалися з використанням фреймворку Optuna, який реалізовує алгоритм Tree-structured Parzen Estimator (TPE) [6, 7]. Цей підхід дозволив ефективно дослідити простір параметрів і підібрати оптимальні конфігурації шляхом мінімізації середньоквадратичної помилки

(MSE) на валідаційному наборі в умовах 5-кратної крос-валідації. Завдяки цьому забезпечувалося підвищення точності моделювання й надійність керування мікрокліматом у реальних умовах [3].

Моделі прогнозування мікрокліматичних параметрів у великих приміщеннях базуються на математичних функціях, що описують залежності між вхідними змінними та очікуваними результатами. Зокрема, у випадку множинної лінійної регресії використовується функція, яка встановлює прямий зв'язок між вхідними параметрами та прогнозованими значеннями. У моделі опорних векторів формалізується задача пошуку оптимальної межі для виявлення складних залежностей у даних, тоді як штучна нейронна мережа відтворює багаторівневу обробку сигналів, що проходять через вхідний, прихований і вихідний шари, забезпечуючи моделювання складної динаміки змін мікроклімату.

Модель множинної лінійної регресії (MLR) у цьому дослідженні визначала коефіцієнти за методом найменших квадратів, мінімізуючи суму квадратів похибок між прогнозованими та фактичними значеннями. Така структура дозволяє оцінити внесок кожної незалежної змінної у формування вихідної, спираючись на лінійну залежність без врахування взаємодій або нелінійностей між параметрами. Попри обмеження щодо моделювання складних зв'язків, MLR залишається інтерпретованим підходом, придатним для аналізу впливу окремих змінних на мікроклімат великого приміщення.

Для реалізації методу опорних векторів (SVM) застосовувалося RBF-ядро, що трансформувало вхідні дані у простір вищої розмірності, утворюючи модель регресії (SVR) для апроксимації нелінійних залежностей. Оптимізація здійснювалась через алгоритм SMO, а для підвищення стійкості до малих похибок використовувалась ϵ -інсенситивна функція втрат, що дозволяє ігнорувати незначні відхилення. Така конфігурація забезпечила точне узагальнення даних і адаптацію до складних змін параметрів середовища.

Штучна нейронна мережа (ANN), що використовувалась у роботі, складалася з вхідного шару для 11 змінних, одного прихованого шару зі 100 нейронами та вихідного шару з одним нейроном, відповідальним за прогноз. Передача сигналів реалізовувалась за допомогою вагових коефіцієнтів, а функція активації ReLU покращувала ефективність навчання, лінійна функція на виході забезпечувала безпосередній розрахунок результатів. Навчання виконувалося алгоритмом Adam із функцією втрат MSE, при цьому частину даних зарезервовано для валідації, а тренування проводилося упродовж 100 епох із пакетом у 10 зразків [5].

Модель XGBoost, заснована на градієнтному бустингу, поєднує функцію втрат і регуляризацию для контролю якості навчання та уникнення перенавчання. Вона поступово вдосконалює передбачення шляхом побудови ансамблю дерев рішень, де кожне наступне дерево навчається на залишкових помилках попередніх. Завдяки послідовному уточненню прогнозів через T ітерацій модель досягає високої точності навіть за обмежених обчислювальних ресурсів, залишаючись масштабованим і ефективним інструментом для керування складними системами.

Фінальне прогнозоване значення визначається як сума прогнозів усіх дерев, помножених на оптимальні вагові коефіцієнти, встановлені під час навчання (рис. 1).

Гіперпараметри моделі XGBoost були підібрані з урахуванням досягнення максимальної продуктивності. Початкове значення швидкості навчання встановлено на рівні 0.1, що знижувало ризик перенавчання, забезпечуючи поступове вдосконалення моделі. Максимальна глибина дерев обмежувалась п'ятьма рівнями, що дозволяло зберігати баланс між точністю прогнозування та здатністю до узагальнення, а кількість дерев дорівнювала 100, що забезпечувало охоплення різноманітних патернів у даних.

Завдяки такій конфігурації модель XGBoost оптимізувала прогноз із мінімальною похибкою та високою стабільністю, що є ключовим для регулювання мікроклімату у великих приміщеннях. Модель дозволяла виконувати передбачення температури, вологості та рівня CO_2 в режимі реального часу, реагуючи на змінні умови середовища. Її здатність до самонавчання та адаптації робить її дієвим інструментом у практичних умовах керування технічними системами.

У процесі навчання XGBoost застосовував механізм градієнтного бустингу, де кожне дерево рішень (CART) будувалося на помилках попередніх. Алгоритм формував послідовність дерев, які на кожному кроці враховували недоліки попереднього прогнозу, поступово знижуючи функцію втрат. Кінцевий прогноз формувався як зважена сума внесків усіх дерев, причому ваги визначалися у процесі оптимізації. Такий підхід, підкріплений регуляризациєю, забезпечував не лише високу точність, а й низьку ймовірність перенавчання.

Завдяки структурі, яка дозволяє ефективно моделювати складні залежності між параметрами, XGBoost демонструє високу ефективність у задачах прогнозування мікроклімату. Вона придатна до використання в умовах змінних навантажень, сезонних коливань або нестабільного зовнішнього середовища, забезпечуючи точні рішення для підтримки стабільних умов у великих приміщеннях [9].

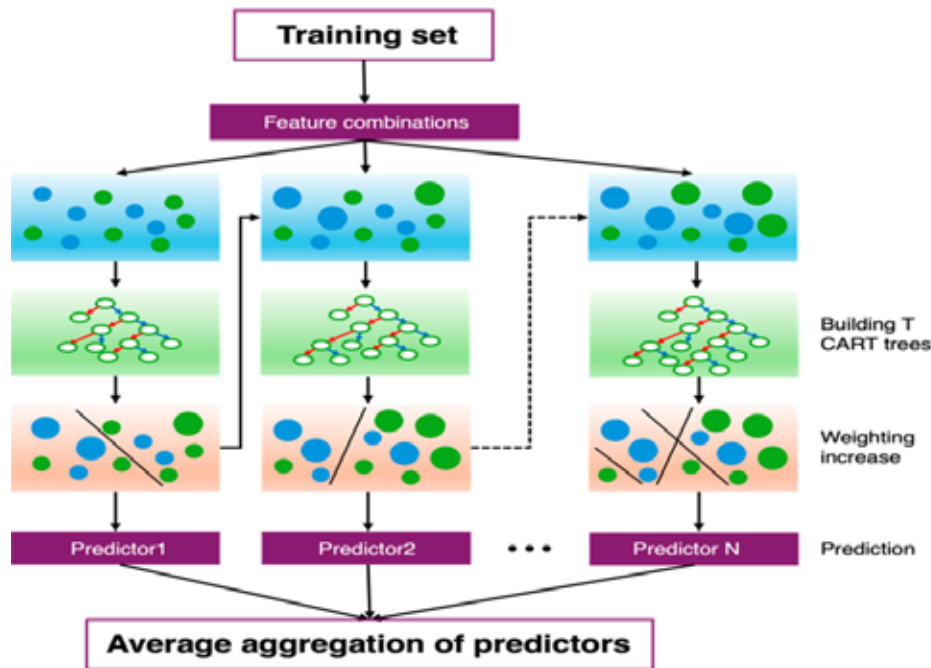


Рис. 1. Схематичне зображення моделі XGBoost

Для оцінки продуктивності моделей, побудованих у цьому дослідженні, було розраховано середньоквадратичну похибку $RMSE$, коефіцієнт детермінації R^2 та залишкову передбачувальну девіацію RPD як для навчального, так і для тестового наборів даних. Ці метрики широко використовуються для оцінювання точності регресійних моделей [13].

Середньоквадратична похибка $RMSE$ (1) визначається як квадратний корінь із середнього значення квадратів різниць між фактичними та прогнозованими значеннями.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

У цьому рівнянні n позначає кількість спостережень, y_i – фактичні значення, $\hat{y}_i$ – прогнозовані значення моделі, а $\bar{y}$ – середнє значення фактичних спостережень.

$RMSE$ використовується для оцінки масштабу похибки, де нижчі значення вказують на кращу точність прогнозування [1]. Цей показник дозволяє оцінити абсолютну помилку моделі та порівняти її ефективність із альтернативними підходами.

Коефіцієнт детермінації R^2 (2) є важливим показником якості моделі, який демонструє, наскільки добре модель пояснює варіабельність даних. Значення R^2 змінюється у діапазоні від 0 до 1. Чим ближче значення R^2 до 1, тим краще модель відображає закономірності у даних і пояснює їх варіативність. Навпаки, значення, близькі до 0, вказують на слабку пояснювальну здатність моделі та високу ступінь невизначеності у прогнозах [4].

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (2)$$

Високе значення R^2 свідчить про те, що модель успішно враховує зв'язки між вхідними параметрами та прогнозованими значеннями, що є критично важливим для точного керування мікрокліматом у великих приміщеннях. Однак сам коефіцієнт детермінації не дає повної картини, тому його слід аналізувати у поєднанні з іншими метриками, такими як $RMSE$ та RPD , для отримання повної оцінки продуктивності моделі.

Залишкова передбачувальна девіація RPD (3) визначається як відношення стандартного відхилення фактичних значень до стандартного відхилення помилок моделі. Ця метрика відображає здатність моделі точно передбачати варіативність даних.

$$RPD = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}{RMSE}. \quad (3)$$

Вищі значення RPD вказують на кращу прогностичну ефективність моделі. Ця метрика є особливо важливою для оцінки ефективності моделей, застосованих у прогнозуванні мікрокліматичних умов у великих приміщеннях.

При проведенні оцінювання ефективності чотирьох моделей штучного інтелекту MLR, SVM, ANN та XGBoost для прогнозування внутрішньої температури $TMPin$, відносної вологості RHM та концентрації вуглекислого газу (CO_2) з горизонтом у 30 хвилин, продуктивність моделей визначалась за допомогою метрик $RMSE$ та R^2 , окремо для навчальної та тестової вибірок, відповідно позначених як $RMSET$, $RMSEP$, RT^2 та RP^2 . Додатково застосовувався індекс залишкової передбачуваної девіації RPD на тестовому наборі, що дозволяє оцінити надійність передбачень, особливо в умовах динамічних змін мікроклімату [9].

Отримані результати дозволяють зробити наступні висновки. XGBoost стабільно демонструвала найкращі показники: для параметру $TMPin$ модель досягла $RMSE$ на найнижчому рівні та найвищого коефіцієнта детермінації – 0.9753 на навчальній і 0.9724 на тестовій вибірках. У випадку з RHM показники були подібними – R^2 відповідно 0.9682 та 0.9656. Щодо прогнозування концентрації CO_2 , XGBoost також перевершила інші моделі, досягнувши найнижчих $RMSE$ (8.5519 і 9.519) та найвищих значень R^2 (0.9940 і 0.9929), що засвідчує її точність і стабільність. ANN у завданні прогнозування росту динь забезпечила $R^2 = 0.9$, а RNN-LSTM для температури теплиці на 30 хвилин уперед – $R^2 = 0.96$. Навіть модель GBR-ET не змогла перевершити XGBoost у точності прогнозів. Значення RPD додатково підтвердили вищу надійність XGBoost. Хоча інші моделі (MLR, SVM, ANN) також показали хороші результати з $RPD > 3.8$, XGBoost продемонструвала істотну перевагу: 6.0197 для $TMPin$, 5.3878 для RHM й 11.8464 для CO_2 . Такі показники підкреслюють виняткову точність і адаптивність моделі для прогнозування мікрокліматичних параметрів у великих приміщеннях, що є ключовим чинником для ефективного управління відповідними технологічними процесами [15].

Додатковий аналіз моделі XGBoost, заснований на графічному зіставленні фактичних та прогнозованих значень (рис. 2) показав, що чим ближче передбачені значення до лінії 1:1, тим точніше модель здатна прогнозувати параметри мікроклімату. Візуальне порівняння дозволяє зробити висновок про стабільність та точність XGBoost у передбаченні ключових екологічних параметрів, що ще раз підтверджує її перевагу над іншими підходами.

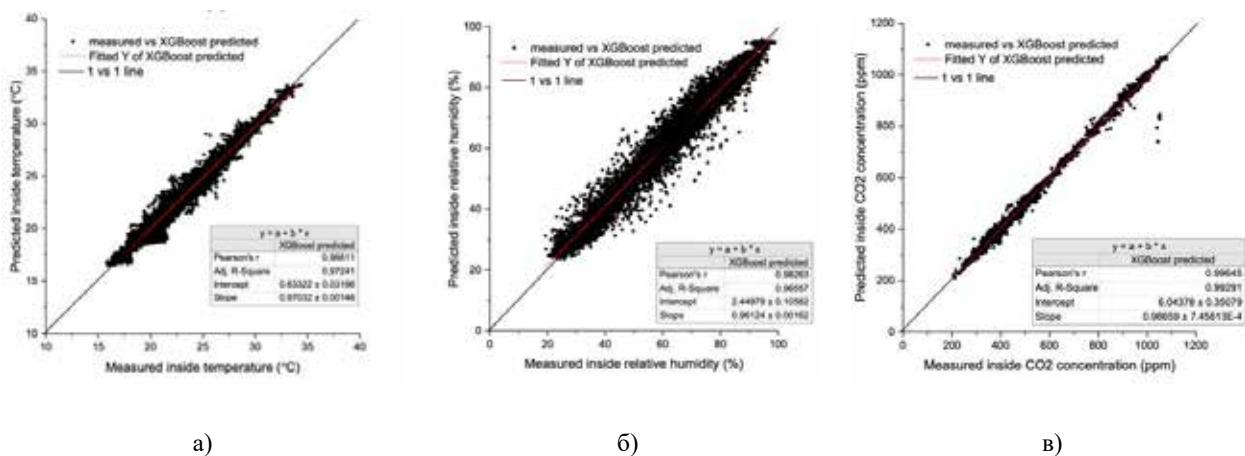


Рис. 2. Порівняння прогнозованих та фактичних значень різних параметрів для моделі XGBoost

У рамках дослідження було проаналізовано ефективність моделі XGBoost у прогнозуванні ключових параметрів – температури, вологості та концентрації CO_2 – з акцентом на виявлення найбільш впливових факторів для кожного з них.

Модель XGBoost продемонструвала найвищу точність та стабільність при прогнозуванні внутрішньої температури на 30 хвилин уперед. Найвпливовішим фактором виявилася поточна температура ($F = 778$), яка є основою для короткострокового передбачення. Другим за значущістю стала сонячна

радіація (SRD, $F = 533$), яка визначає ступінь теплового впливу на приміщення. Час доби *DATETIME*, наявність опадів *RAIN*, зовнішня температура *TMPout* та накопичена сонячна енергія *CSR* також мали суттєвий внесок ($F > 240$), що свідчить про їхню непряму участь у формуванні температури через зміну теплового балансу. Натомість обігрів *HEAT*, напрямок вітру *WDR* та вологість *RHM* мали низькі значення важливості (від 3 до 31), отже, незначно впливали на зміну температури у межах 30 хвилин. Дерево рішень XGBoost підтверджує ці висновки: *TMPin* – ключовий вузол дерева, а *SRD* і *DATETIME* з'являються на глибших рівнях, що вказує на їхній внесок у приріст температури (рис. 3). Ці закономірності підтверджують, що модель не лише стабільна, а й здатна точно ідентифікувати провідні фізичні чинники [8].

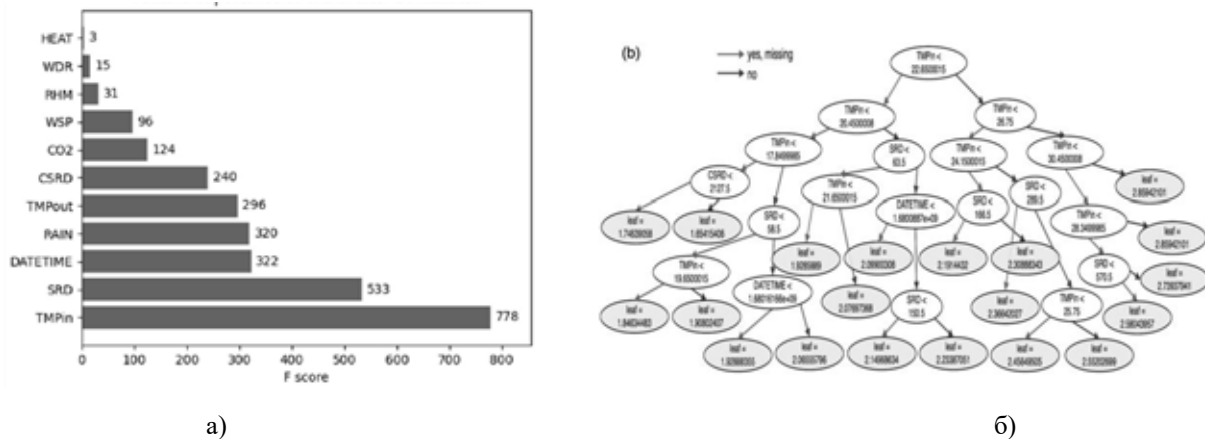


Рис. 3. Аналіз впливу параметрів та структури рішення моделі XGBoost для прогнозування температури

Для прогнозування відносної вологості на аналогічний проміжок модель XGBoost також показала стабільні результати, з найвищим F -скором у параметра *RHM* = 776. Поточний рівень вологості виступає основним предиктором, що узгоджується з інерційною природою цього показника. *SRD* ($F = 512$) виявилася другим чинником, адже вона опосередковано впливає на випаровування води, змінюючи вологість повітря. На відміну від моделі температури, обігріву *HEAT* мав помітний вплив на вологість, адже обігріте повітря сприяє висушуванню середовища. Також вагомими залишаються *TMPin* й *DATETIME*, що підтверджує комплексну залежність вологості від кількох умов. Дерево рішень для вологості має складнішу структуру та більшу глибину, ніж для температури, що свідчить про багатофакторну природу вологості (рис. 4). Таким чином, XGBoost показує не лише високу точність, а й гнучкість у відтворенні взаємодії між параметрами.

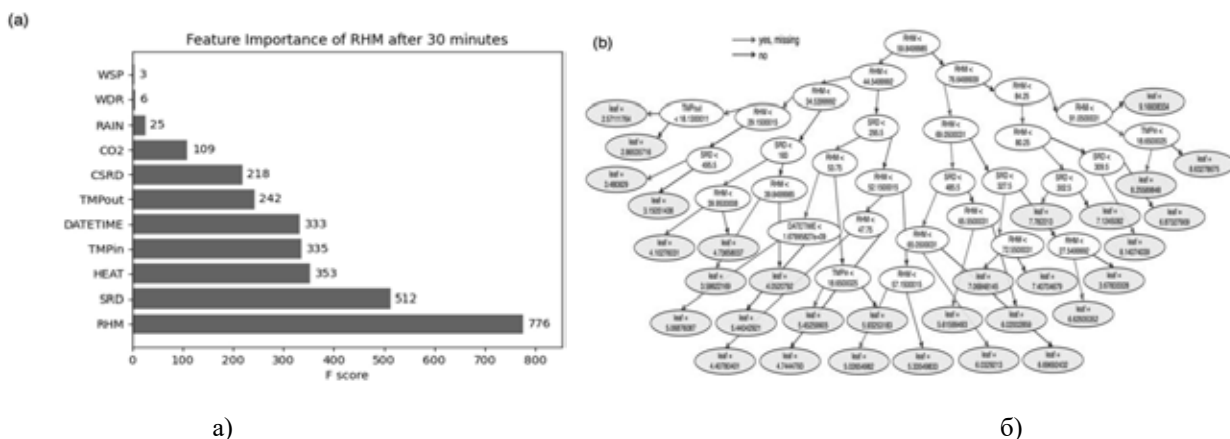
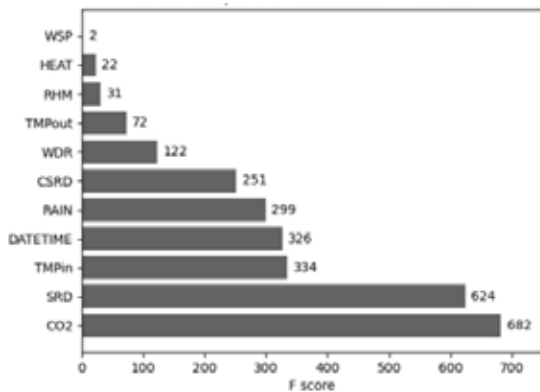


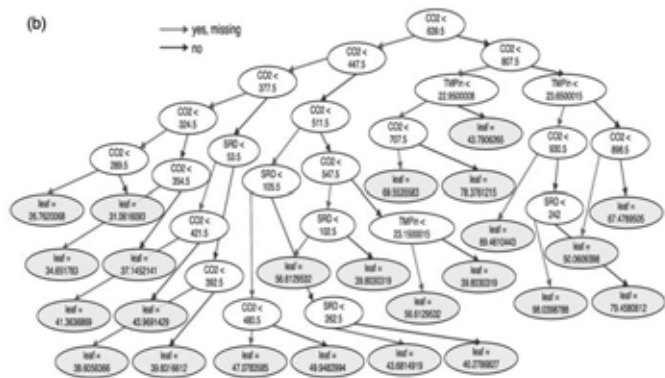
Рис. 4. Аналіз впливу параметрів та структури рішення моделі XGBoost для прогнозування відносної вологості

Прогнозування CO_2 також спирається на поточне значення цього параметра, яке є провідним ($F = 682$). Це зумовлено повільною дифузією газу та прямим зв'язком між попереднім і майбутнім станами середовища (рис. 5). Водночас SRD ($F = 624$) має важливий вплив, адже вона стимулює фотосинтетичну активність рослин, які поглинають CO_2 , особливо в денні години.

Додатковими чинниками стали температура (TMPin), час (DATETIME) та опади (RAIN), кожен із яких має значення понад 299. Їх внесок зумовлений тим, що вони впливають на повітрообмін, фотосинтез та вологість – тобто, на біологічну динаміку вмісту CO_2 .



a)



б)

Рис. 5. Аналіз впливу параметрів та структури рішення моделі XGBoost для прогнозування відносної вологості

Модель XGBoost виокремлює два сценарії: природне зниження CO_2 при зростанні SRD та TMPin , та ще стрімкіше зниження у разі карбонової фертизації, коли рівень газу перевищує 800 ppm. Структура дерева підтверджує здатність моделі враховувати як природні, так і штучно індуковані процеси. Це доводить ефективність XGBoost у системах, що потребують точного контролю за рівнем CO_2 , наприклад, у великих офісних приміщеннях [10].

Таким чином, модель XGBoost продемонструвала не лише високу точність прогнозування, а й здатність виявляти основні фізичні та біологічні чинники, що визначають динаміку параметрів мікроклімату у великих приміщеннях.

Висновки. У цьому дослідженні було розглянуто моделі машинного навчання для прогнозування мікроклімату у великих приміщеннях та досліджено їх ефективність. Порівняння продуктивності моделей за метриками точності показало, що XGBoost продемонструвала найкращі результати, маючи найнижчі значення $RMSE$ серед усіх наборів даних для всіх трьох прогнозованих параметрів. Крім того, вона досягла найвищих значень R^2 : 0.9724 – для температури, 0.9656 – для вологості, 0.9929 для концентрації CO_2 . Значення RPD також підтвердили домінування XGBoost над іншими моделями, особливо у прогнозуванні CO_2 , де параметр RPD досяг 11.8464, що суттєво перевищує результати інших моделей. Аналіз графіків розсіювання (1:1) фактичних і прогнозованих значень для тестового набору показав, що точки XGBoost були найближчими до ідеальної лінії відповідності. Особливо у випадку CO_2 , де високі значення RPD ще раз підтвердили здатність XGBoost ефективно прогнозувати зміну цього параметра. Дослідження важливості параметрів у XGBoost показало, що зміни мікроклімату у приміщенні значною мірою визначаються часом доби та рівнем сонячної радіації. Аналіз першого дерева рішень XGBoost підтвердив, що модель ефективно навчається розпізнавати взаємодію між внутрішніми умовами приміщення, фізіологією культур та діями оператора. Таким чином, застосування XGBoost у керуванні мікрокліматом дозволяє точно прогнозувати температуру, вологість й CO_2 . Якщо передбачувані значення виходять за межі оптимальних, система може автоматично активувати виконавчі механізми для їх коригування.

Крім того, можливість інтерпретувати процес ухвалення рішень моделі дає змогу оператору краще розуміти, як змінюються кліматичні параметри у відповідь на роботу систем управління, що дозволяє приймати більш обґрунтовані рішення щодо керування середовищем.

У цій тематиці гібридні сховища інформації забезпечують ефективне зберігання, обробку та доступ до великих обсягів мікрокліматичних даних у реальному часі. Вони поєднують переваги реляційних та

розподілених баз даних, що дозволяє адаптувати систему до динамічних умов середовища та підвищити стабільність роботи моделей прогнозування.

Попри високу точність моделі, набір даних, використаний у дослідженні, не охоплював усі можливі фактори, що впливають на мікроклімат. Через це у деяких випадках прогнозовані значення відрізнялися від фактичних.

Список використаних джерел:

1. Agarap A. F. Deep Learning Using Rectified Linear Units (ReLU). arXiv preprint. 2019. arXiv:1803.08375.
2. Agatonovic-Kustrin S., Beresford R. Basic Concepts of Artificial Neural Network (ANN) Modeling and Its Application in Pharmaceutical Research. *J. Pharm. Biomed. A.* 2000. Vol. 22. P. 717–727. DOI: [https://doi.org/10.1016/S0731-7085\(99\)00272-1](https://doi.org/10.1016/S0731-7085(99)00272-1).
3. Akiba T. et al. Optuna: A Next-Generation Hyperparameter Optimization Framework. Proc. 25th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining. Anchorage, AK, USA, 2019. P. 2623–2631. DOI: <https://doi.org/10.1145/3292500.3330701>.
4. AL ISSA H. A. Regulating steady-state voltage deviation using fuzzy logic / Huthaifa AL Issa, L. Hussienat, A. Panov, K. Demchenko, O. Piskarov, O. Miroshnyk, T. Shchur // *Przegląd Elektrotechniczny*. 2025. T. 2025. № 3.
5. Awad M., Khanna R. Support Vector Regression. In: *Efficient Learning Machines*. – Berkeley: Apress, 2015. P. 67–80. ISBN: 978-1-4302-5990-9.
6. Bentéjac C., Csörgő A., Martínez-Muñoz G. A Comparative Analysis of Gradient Boosting Algorithms. *Artif. Intell. Rev.* 2021. Vol. 54. P. 1937–1967. DOI: <https://doi.org/10.1007/s10462-020-09896-5>.
7. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining. San Francisco, CA, USA, 13–17 Aug. 2016. P. 785–794. DOI: <https://doi.org/10.1145/2939672.2939785>.
8. Chicco D., Warrens M. J., Jurman G. The Coefficient of Determination R^2 Is More Informative than SMAPE, MAE, MAPE, MSE and RMSE. *PeerJ Comput. Sci.* 2021. Vol. 7. e623. DOI: <https://doi.org/10.7717/peerj-cs.623>.
9. Hodson T. O. Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not. *Geosci. Model Dev.* 2022. Vol. 15. P. 5481–5487. DOI: <https://doi.org/10.5194/gmd-15-5481-2022>.
10. Karajan M., Sorenson P., Saleh K. Hybrid Cloud Storage: Research Status, Challenges, and Future Directions. *Journal of Cloud Computing: Advances, Systems and Applications*. 2020. Vol. 9. Art. 49. DOI: <https://doi.org/10.1186/s13677-020-00188-y>.
11. Nguyen N.-H. et al. Efficient Estimating Compressive Strength of Ultra-High Performance Concrete Using XGBoost Model. *J. Build. Eng.* 2022. Vol. 52. Art. 104302. DOI: <https://doi.org/10.1016/j.jobbe.2022.104302>.
12. Rittinghouse J. W., Ransome J. F. *Cloud Computing: Implementation, Management, and Security*. 2nd ed. Boca Raton: CRC Press, 2017. ISBN 978-1-4987-3950-2.
13. Salem H. et al. Predictive Modelling for Solar Power-Driven Hybrid Desalination System Using ANN with Adam Optimization. *Desalination*. 2022. Vol. 522. Art. 115411. DOI: <https://doi.org/10.1016/j.desal.2021.115411>.
14. Taki M. et al. Applied Machine Learning in Greenhouse Simulation; New Application and Analysis. *Inf. Process. Agric.* 2018. Vol. 5. P. 253–268. DOI: <https://doi.org/10.1016/j.inpa.2018.03.003>.
15. Tymchuk S., Abramenko I., Shendryk V., Shendryk S., Piskarev O. Controller Software Optimization in Adaptive Extreme Automation Systems. *Lecture Notes in Networks and Systems*. 2022. Vol. 472. P. 252–259. ISBN 978-3-031-05229-3.

Дата надходження статті: 25.06.2025

Дата прийняття статті: 01.07.2025

Опубліковано: 23.09.2025