

ДОРОШЕНКО ДМИТРО ВОЛОДИМИРОВИЧ ✉ – здобувач кафедри передачі електричної енергії, Національний технічний університет «Харківський політехнічний інститут»; м. Харків, Україна; e-mail: ddorosenko55@gmail.com.

ПРОГНОЗУВАННЯ ВИКИДІВ ДІОКСИДУ ВУГЛЕЦЮ (CO_2) З ГАЗОТУРБІННИХ УСТАНОВОК ЗА ДОПОМОГОЮ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ

У роботі розглядається застосування методів машинного навчання, зокрема моделей XGBoost та Random Forest, для прогнозування рівня викидів вуглецю оксиду (CO_2) з газотурбінного обладнання. Точне передбачення рівня шкідливих викидів є критично важливим завданням у контексті забезпечення енергетичної безпеки та досягнення екологічної стійкості. У дослідженні використано реальний набір даних, зібраний у Туреччині упродовж 2011–2015 років, що містить понад 36 тисяч записів щогодинного моніторингу технічних та атмосферних параметрів газотурбінної установки. З метою покращення якості моделювання було проведено попередню обробку даних, що включала очищення, аналіз кореляції ознак та стандартизацію числових значень. На основі обраних параметрів побудовано дві регресійні моделі: XGBRegressor та RandomForestRegressor, які продемонстрували високу точність при оцінюванні результатів. Найкраще значення коефіцієнта детермінації склало 0,742 для моделі Random Forest, що вказує на її ефективність у відтворенні залежностей між технічними параметрами установки та рівнем викидів CO_2 . Аналіз важливості ознак показав, що найбільший вплив на рівень емісій мають параметри, пов'язані з температурними режимами турбіни, тоді як зовнішні атмосферні умови відіграють другорядну роль. Отримані результати мають практичне значення для енергетичної галузі, оскільки можуть бути використані для моніторингу викидів, планування модернізації енергетичних об'єктів, обґрунтування екологічної політики та вибору оптимальних режимів експлуатації устаткування. Впровадження подібних моделей у промислову практику сприятиме зменшенню викидів парникових газів, підвищенню ефективності енергоспоживання та реалізації національних стратегій сталого розвитку. Дослідження підтверджує, що застосування сучасних аналітичних інструментів на основі машинного навчання може зробити вагомий внесок у забезпечення екологічної та енергетичної безпеки країни.

Ключові слова: прогнозування викидів CO_2 ; енергетична безпека; машинне навчання; Random Forest; XGBoost; регресійна модель; газотурбінна установка; стандартизація даних.

Вступ. Зменшення викидів парникових газів, зокрема CO_2 , є ключовим завданням для сталого розвитку і енергетичної безпеки. CO_2 – це основний парниковий газ, що спричиняє глобальне потепління; тому точне прогнозування його емісій допомагає планувати заходи з їх скорочення і розвитку «зеленої» енергетики. З іншого боку, зниження викидів CO_2 пов'язане з диверсифікацією енергетичних ресурсів і зменшенням залежності від імпорту викопного палива, що підвищує енергетичну безпеку країни.

Газотурбінні установки широко використовуються в електроенергетиці завдяки своїй високій ефективності (до 60 % у комбінованому циклі) та здатності швидко реагувати на зміни попиту, що робить їх ідеальними для балансування навантаження в умовах нестабільності відновлюваних джерел енергії.

Двигун силової установки (газова турбіна) в основному використовується для вироблення електроенергії. Крім того, порівняно з вугільними ТЕС такі установки сприяють значному зниженню викидів – до 65 % меншого обсягу CO_2 на одиницю виробленої електроенергії [1, 2].

Мета статті. Метою даної роботи є розробка та оцінка регресійних моделей машинного навчання (моделей Random Forest та XGBoost) для прогнозування викидів (вуглецю оксиду) з турбінних установок на основі набору даних, а також аналіз отриманих результатів з погляду їх значущості для енергетичної безпеки.

Літературний огляд. Питання прогнозування викидів CO_2 є надзвичайно важливим з огляду на зміну клімату та екологічне регулювання, тому застосовуються різні методи машинного навчання. Так, у дослідженні Mădălina Maria Muraru та ін. (2025)

проаналізували автопарки понад 26 500 транспортних засобів, порівнявши Random Forest, XGBoost і нейронні мережі, і встановили, що Random Forest забезпечує найкращу точність CO_2 -прогнозів [3]. У дослідженні Jabeur та ін. (2024) виявили, що ансамблеві моделі, зокрема Random Forest і XGBoost, значно перевершують традиційні регресійні підходи при моделюванні екологічних викидів [4]. У роботі Hou & Liu (2024), яка використовує XGBoost із байєсівською оптимізацією (BO-XGBoost) для прогнозування викидів CO_2 у будівельній галузі Китаю, досягли високих результатів, підкресливши ефективність підходу [5]. Аналогічно Finger's група зафіксувала, що підтримка RF для селекції ознак плюс BO-XGBoost дуже точно прогнозує рівні парникових газів у розвинених і країнах, що розвиваються [6]. Нарешті, дослідження на сміттєспалювальному заводі показало, що як RF, так і XGBoost успішно моделюють концентрацію CO_2 , використовуючи понад 20 вхідних параметрів [7].

Многу обрані Random Forest та XGBoost через їхню високу точність та стійкість: RF ефективно справляється із зашумленими та категорійними даними, а XGBoost із регуляризациєю й жорстким ансамблем демонструє найкращі результати в багатьох дослідженнях.

Теоретичні основи. Регресія – це підхід у машинному навчанні для прогнозування неперервних числових значень на основі вхідних ознак. На відміну від класифікації, яка відносить об'єкт до одного з дискретних класів, регресійні моделі виводять числові прогнози (наприклад, концентрацію забруднювача) на основі історичних даних. У сучасних рішеннях часто використовують ансамблеві методи на основі дерев

© Д. В. Дорошенко, 20xx



Ця робота ліцензується відповідно до **Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)**
Конфлікт інтересів: Автор заявив про відсутність конфлікту

рішень, які поєднують результати багатьох слабких моделей для отримання стійкішого прогнозу.

Випадковий ліс (Random Forest) – це широко розповсюджений ансамблевий алгоритм, який будується на основі поєднання багатьох незалежних дерев рішень. Кожне дерево навчається на випадковій вибірці даних і випадковому піднаборі ознак. Остаточний прогноз отримується шляхом усереднення прогнозів усіх дерев. Такий підхід знижує ризик перенавчання і дає більш надійні прогнози. Random Forest добре працює як для задач класифікації, так і регресії. Окрім високої точності, одним із важливих плюсів Random Forest є те, що він може оцінювати відносну важливість ознак у моделі. У контексті регресії алгоритм Random Forest агрегує результати дерев, що робить модель стійкою до шуму даних. У дослідженнях з прогнозування викидів вуглецю відзначено, що Random Forest дає зрозумілі результати і корисний для виявлення ключових факторів впливу на емісії [8–10]. Для задач регресії:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N f_i(x); \quad (1)$$

де $\hat{y}$ – фінальний прогноз моделі (середнє значення);
 N – кількість дерев у лісі;

$f_i(x)$ – прогноз i -го дерева для вхідного вектора x .

XGBoost (Extreme Gradient Boosting) – це ефективна реалізація градієнтного бустингу дерев рішень. Назва «Extreme Gradient Boosting» підкреслює приналежність моделі до сімейства градієнтних бустингових машин. Градієнтний бустинг – це послідовне побудова дерев, де кожне нове дерево навчається коригувати помилки попередніх. XGBoost оптимізовано для швидкого навчання на великих даних, має регуляризацию для контролю складності моделі і можливості обробки пропусків. Базовий принцип XGBoost можна коротко сформулювати як мінімізацію цільової функції, що включає функцію втрат і регуляризацийний член:

$$\mathcal{L}^{(t)} = \sum_{i=1}^N l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t); \quad (2)$$

де l – функція втрат;

$\hat{y}_i^{(t-1)}$ – прогноз моделі після $t-1$ кроків;

$f_i(x)$ – прогноз i -го дерева для вхідного вектора x ;

$f_t(x_i)$ – прогноз нового дерева;

$\Omega(f_t)$ – регуляризацийний термін, який контролює складність дерева.

Регуляризацийна функція має вигляд:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (2)$$

де T – кількість листків дерева;

w_j – ваги в листках;

γ – штраф за кількість листків (L0-регуляризація);

λ – коефіцієнт L2-регуляризації.

Базовий принцип XGBoost можна коротко сформулювати так: модель мінімізує функцію якості (наприклад, MSE для регресії) плюс регуляризацийний штраф за складність дерев. XGBoost широко застосовується для прогнозування часових рядів і емісій, оскільки він здатний захоплювати нелінійні залежності та взаємодію ознак. У численних дослідженнях XGBoost демонстрував високу точність прогнозів CO₂ та інших забруднювачів – він часто перевершує інші моделі з погляду помилки прогнозу. Водночас XGBoost потребує налаштування гіперпараметрів та більших обчислювальних ресурсів [11–13].

Обидва підходи – Random Forest і XGBoost – належать до ансамблевих дерев'яних методів. Random Forest використовує «випадковий підсув» (bagging) даних і ознак для побудови незалежних дерев, у той час як XGBoost поєднує (сумує) багато дерев послідовно, кожне з яких вчиться покращувати похибку попереднього. Обидва алгоритми добре справляються з багатовимірними вхідними даними і здатні моделювати складні, нелінійні закономірності, що є ключовим при аналізі викидів залежно від різних факторів.

Опис даних та їх підготовка. Для реалізації задач прогнозування викидів оксиду вуглецю, що є індикатором рівня емісій CO₂ у процесах спалювання палива, у роботі було використано реальний набір даних із відкритого джерела – Gas Turbine CO and NO_x Emission Dataset з репозиторію машинного навчання UCI [14]. Цей датасет широко застосовується в дослідженнях, присвячених екологічному моніторингу, енергетичному моделюванню та аналізу експлуатаційних режимів промислових турбін. Дані були зібрані в рамках щогодинного моніторингу роботи газотурбінної електростанції в Туреччині в період з 2011 по 2015 роки, що забезпечує достатній часовий горизонт для формування надійних прогнозних моделей. Загальний обсяг даних становить 36733 спостереження. Кожен запис містить 11 числових ознак, що описують як внутрішні параметри турбінного агрегату, так і зовнішні атмосферні умови. До ключових змінних відносяться:

1. AP – зовнішній тиск повітря (мбар).
2. AH – відносна вологість повітря (%).
3. AFDP (Air Filter Difference Pressure) – різниця тиску на повітряному фільтрі (мбар).
4. GTER (Gas Turbine Exhaust Pressure) – тиск на виході турбіни (мбар).
5. TIT (Turbine Inlet Temperature) – температура на вході в турбіну (°C).
6. TAT (Turbine After Temperature) – температура на виході з турбіни (°C).
7. TEY (Turbine Energy Yield) – енергетичний вихід турбіни (МВт·ч).
8. CDP (Compressor Discharge Pressure) – тиск у компресорі на виході (мбар).
9. CO – концентрація оксиду вуглецю у викидах (мг/м³).

Попередній аналіз даних включав перевірку на пропущені значення, які були відсутні, що дало змогу

використовувати повну вибірку без додаткових процедур заповнення. Для зменшення ризику перенавчання та покращення якості моделей було проведено аналіз кореляції між ознаками та цільовою змінною CO₂ (рис. 1). Ознаки з низьким рівнем кореляції, такі як АН, ТАТ та АР, були виключені з подальшого аналізу.

Оскільки вхідні параметри мають різні фізичні одиниці вимірювання та варіації масштабів, було застосовано стандартизацію даних методом Z-перетворення (StandardScaler) [15], що дозволяє кожній ознаці мати середнє значення, близьке до нуля, та стандартне відхилення, близьке до одиниці.

Це забезпечує рівнозначний внесок кожної змінної у навчання моделей машинного навчання та запобігає домінуванню ознак з великими числовими значеннями, тому стандартизуємо кожне значення за формулою:

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

де x – значення ознаки, що стандартизується;

μ – середнє значення ознаки в навчальному наборі;

σ – стандартне відхилення ознаки в навчальному наборі.

Результатом цієї процедури є рис. 2, в якій усі ознаки мають нульове середнє значення та стандартне відхилення, близьке до одиниці. Такий підхід покращує збіжність алгоритмів градієнтної оптимізації та забезпечує рівний вплив усіх параметрів на функцію втрат моделі.

Далі проведено розподіл підготовлених даних на навчальну та тестову вибірки у співвідношенні 80:20.

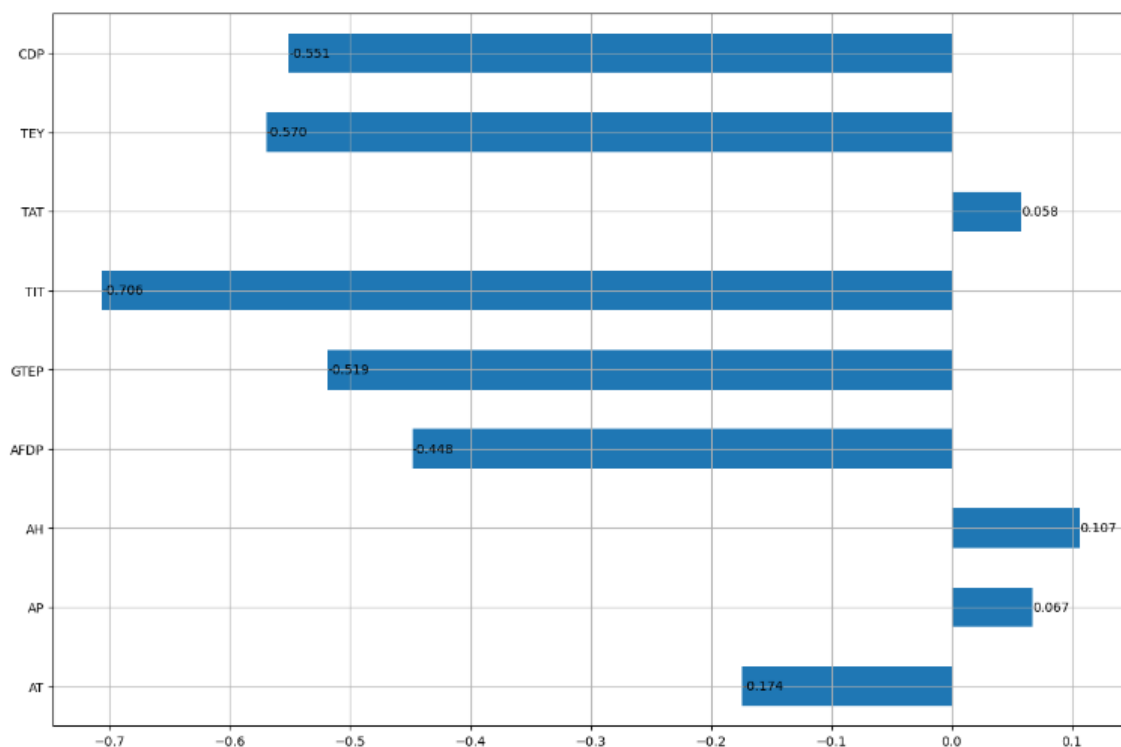


Рисунок 1 – Діаграма кореляції параметрів з CO₂

	AT	AFDP	GTEP	TIT	TEY	CDP
count	3.673300e+04	3.673300e+04	3.673300e+04	3.673300e+04	3.673300e+04	3.673300e+04
mean	-1.176081e-16	-1.015144e-15	5.292366e-16	9.609823e-15	9.888369e-16	-4.704326e-16
std	1.000014e+00	1.000014e+00	1.000014e+00	1.000014e+00	1.000014e+00	1.000014e+00
min	-3.215577e+00	-2.375059e+00	-1.874640e+00	-4.597825e+00	-2.144032e+00	-2.028623e+00
25%	-7.964882e-01	-7.363991e-01	-5.802810e-01	-5.490425e-01	-5.798539e-01	-5.745191e-01
50%	1.185304e-02	1.574091e-02	-1.095835e-01	2.550115e-01	1.431620e-02	-8.773591e-02
75%	7.992472e-01	5.832377e-01	8.334799e-01	8.879903e-01	6.769952e-01	7.296924e-01
max	2.603647e+00	4.761549e+00	3.611191e+00	1.110388e+00	2.944830e+00	2.845821e+00

Рисунок 2 – Застосування StandardScaler та описова статистика даних

Це класичний підхід, який дозволяє навчити модель на більшості даних, а потім оцінити її узагальнюючу здатність на нових, ще не бачених прикладах. Додатково, у ході тренування моделей була використана стратегія фіксованого генератора випадкових чисел (`random_state=42`), що гарантує відтворюваність результатів і дозволяє об'єктивно порівнювати моделі між собою. Усі кроки обробки даних реалізовані з використанням мови програмування Python та бібліотек `pandas` (для обробки таблиць), `matplotlib` (для побудови графіків), `seaborn` (для візуалізації кореляцій), `scikit-learn` (для масштабування та поділу даних), а також `xgboost` і `sklearn.ensemble` для побудови моделей. Загалом, підготовка даних здійснювалась з урахуванням принципів чистоти даних, оптимізації вхідного простору та забезпечення стійкості моделі до варіацій вхідних ознак, що є фундаментом для подальшого ефективного навчання регресійних моделей XGBoost та Random Forest.

Побудова моделей і обробка результатів. XGBoost (Extreme Gradient Boosting) реалізує принцип градієнтного бустингу, за якого кожне наступне дерево коригує помилки попередніх. Для забезпечення оптимального балансу між точністю та стійкістю моделі до перенавчання були підібрані такі основні гіперпараметри:

- `n_estimators = 480` – кількість дерев у ансамблі;
- `learning_rate = 0,1` – темп навчання (менші значення потребують більше дерев, але підвищують узагальнення);
- `max_depth = 3` – обмеження глибини дерев для уникнення перенавчання;
- `min_child_weight = 1` – мінімальна сума ваг у дочірньому вузлі;
- `gamma = 1` – мінімальне зменшення функції втрат для розділення вузлів;
- `subsample = 0,8` та `colsample_bytree = 0,8` – параметри випадкового вибору даних і ознак;
- `objective = 'reg:squarederror'` – цільова функція для регресії;

- `reg_alpha = 1` та `reg_lambda = 1` – L1- і L2-регуляризація відповідно;
- `random_state = 42` – фіксований генератор випадкових чисел для відтворюваності.

Модель була навчена на 80 % даних, а решта 20 % використана для тестування. Оцінка точності моделі здійснювалась за допомогою коефіцієнта детермінації (R^2) [16], який визначає частку варіації цільової змінної, поясненої моделлю:

$$R^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (4)$$

де y_i – фактичне значення;

$\hat{y}_i$ – середнє значення ознаки в навчальному наборі;

$\bar{y}$ – середнє фактичних значень.

Результати розрахунку коефіцієнта детермінації для обох моделей наведені у табл. 1. Результат моделі XGBoost Regressor на тестовій вибірці становив $R^2 = 0,735$, що свідчить про здатність моделі пояснити приблизно 73,5 % варіації у викидах CO_2 . Візуальне порівняння прогнозованих та фактичних значень, представлене у вигляді лінійної діаграми, показало задовільну відповідність трендів, хоча деякі відхилення спостерігаються в пікових значеннях (рис. 3).

Таблиця 1 – Результати розрахунку коефіцієнта детермінації

Модель	R^2 (коефіцієнт детермінації)
XGBoost Regressor	0,735
Random Forest Regressor	0,742

Алгоритм Random Forest реалізує підхід *bagging*, формуючи велику кількість незалежних дерев рішень, побудованих на випадкових підмножинах даних і ознак. Остаточний прогноз є результатом усереднення передбачень усіх дерев. Було обрано такі параметри моделі:

- `n_estimators = 1000` – кількість дерев;
- `max_depth = 15` – максимальна глибина дерева;

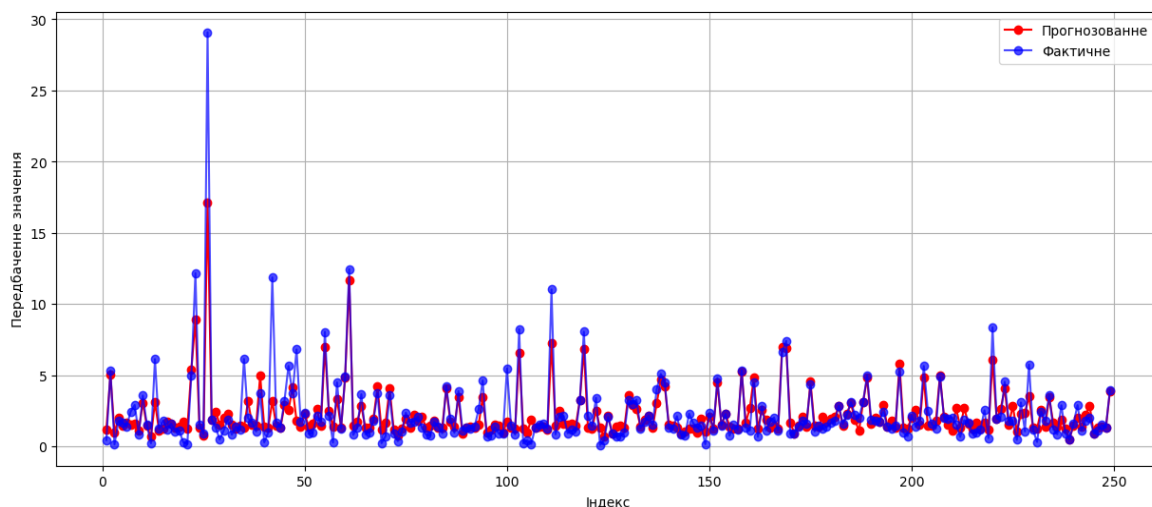


Рисунок 3 – Графік розсіювання передбачених значень для XGBRegressor

- `min_samples_split = 2` – мінімальна кількість зразків для поділу вузла;
- `min_samples_leaf = 1` – мінімальна кількість зразків у листі;
- `max_features = 'log2'` – кількість ознак для поділу;
- `bootstrap = True` – використання бутстреп-зразків;
- `n_jobs = -1` – використання всіх доступних процесорів;
- `random_state = 42`.

Після навчання модель Random Forest продемонструвала $R^2 = 0,742$, що є дещо вищим за показник XGBoost. Це свідчить про кращу здатність моделі відтворювати залежність між технічними параметрами турбіни та рівнем емісій CO₂. Візуалізація також підтвердила більш точне прилягання до фактичних значень у порівнянні з попередньою моделлю (рис. 4).

Висновки. Проведений аналіз продемонстрував, що методи машинного навчання на базі дерев (Random Forest та XGBoost) здатні точно моделювати залежність викидів окису вуглецю від змінних стану газотурбінних установок. Отримані моделі показали високе значення R^2 ($\approx 0,74$) і низьку помилку, що підтверджує їх придатність для задачі прогнозування викидів CO₂ (окису вуглецю). Найважливішими факторами виявилися внутрішні параметри турбіни (зокрема, температури TIT, TAT), а зовнішні умови відіграють другорядну роль. Це узгоджується з очікуваннями: параметри роботи агрегату безпосередньо визначають інтенсивність спалювання палива і утворення шкідливих речовин.

З точки зору енергетичної безпеки, прогнози викидів CO₂ важливі для планування енергетичної стратегії та екологічної політики. Такі моделі можуть допомогти оцінювати вплив різних режимів роботи обладнання чи заміни палива (наприклад, перехід від вугілля на газ) на сумарні викиди. Зменшення викидів вуглецю у енергетиці сприяє скороченню залежності від «брудних» видів палива і підвищенню

енергоефективності, що безпосередньо зміцнює енергетичну безпеку країни. Прогнозуючи майбутні рівні емісій, можна обґрунтовано планувати «зелену» модернізацію генерації, інвестувати в нові технології та оцінювати доцільність політичних рішень (наприклад, стимулювання ВДЕ або перехід на менш вуглецеве паливо) з урахуванням кліматичних цілей.

Отже, виконане дослідження демонструє ефективність регресійних моделей XGBoost і Random Forest для прогнозування викидів CO₂ у контексті енергетичної безпеки. Отримані результати можна використовувати для аналізу енергетичних систем і моніторингу викидів, що важливо для сталого розвитку енергетики і досягнення екологічних цілей.

Список літератури

1. Комітет Верховної Ради України з питань екологічної політики та природокористування. Інформаційно-аналітична записка щодо стану законодавства з питань зміни клімату та основних напрямків діяльності Верховної Ради України у цій сфері. 2020. 5 с. URL: <https://komekolog.rada.gov.ua/uploads/documents/35959.pdf>.
2. What is a gas turbine power station?. *Allied Power Group*. URL: <https://alliedpg.com/latest-articles/gas-turbine-power-station/>.
3. Muraru M. M., Simó Z., Iantovics L. B. Study on state-of-the-art machine learning algorithms efficiency for CO₂ emissions prediction in fleet management optimization. *The 18th International Conference Interdisciplinarity in Engineering (Inter-Eng 2024)*: Lecture Notes in Networks and Systems, vol. 1249, Targu Mures, Romania, 3–4 October 2024. Cham, 2025. P. 1–11. DOI: https://doi.org/10.1007/978-3-031-81685-7_1.
4. Han T. T. T., Lin C.-Y. Exploring long-run CO₂ emission patterns and the environmental kuznets curve with machine learning methods. *Innovation and green development*. 2025. Vol. 4, no. 1. 100195. DOI: <https://doi.org/10.1016/j.igd.2024.100195>.
5. Hou Y., Liu S. Predictive modeling and validation of carbon emissions from China's coastal construction industry: a BO-XGBoost ensemble approach. *Sustainability*. 2024. Vol. 16, no. 10. 4215. DOI: <https://doi.org/10.3390/su16104215>.
6. Gan N., Zhao S. Global greenhouse gas reduction forecasting via machine learning model in the scenario of energy transition. *Journal of Environmental Management*. 2024. Vol. 371. 123309. DOI: <https://doi.org/10.1016/j.jenvman.2024.123309>.
7. Machine learning-based prediction of the CO₂ concentration in the flue gas and carbon emissions from a waste incineration plant / Y. Ma et al. *ACS ES&T Engineering*. 2024. Vol. 4, no. 3. P. 737–747. DOI: <https://doi.org/10.1021/acsestengg.3c00461>.

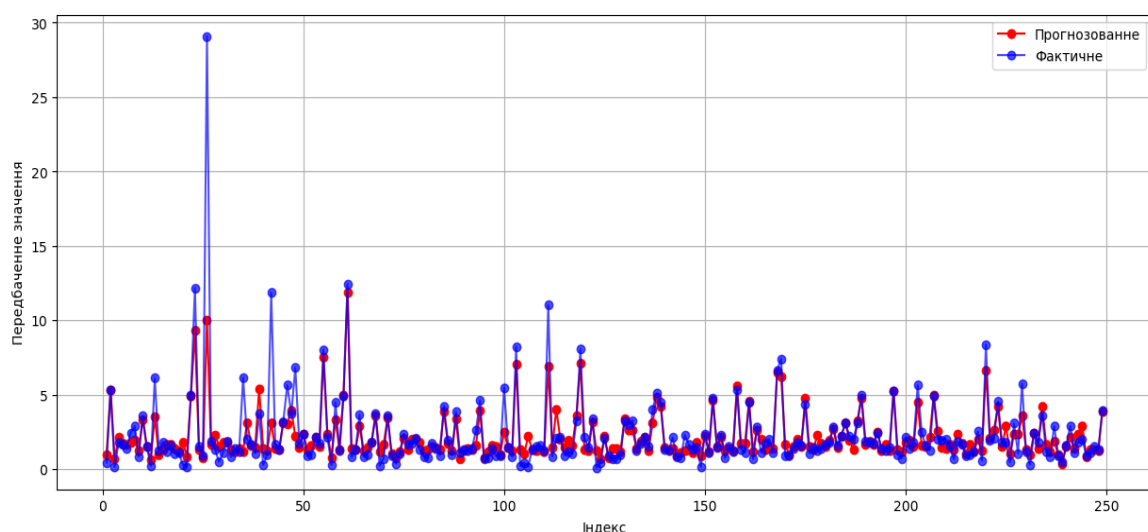


Рисунок 4 – Графік розсіювання передбачених значень для RandomForestRegressor

8. Biau G., Scornet E. A Random Forest guided tour. 2015. 42 p. (Preprint. arXiv:1511.05741). DOI: <https://doi.org/10.48550/arXiv.1511.05741>.
9. Random forest algorithm in machine learning. *GeeksforGeeks*. URL: <https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/>.
10. A comparison of random forest variable selection methods for classification prediction modeling / J. L. Speiser et al. *Expert Systems with Applications*. 2019. Vol. 134. P. 93–101. DOI: <https://doi.org/10.1016/j.eswa.2019.05.028>.
11. Chen T., Guestrin C. XGBoost: a scalable tree boosting system. 2016. 13 p. (Preprint. arXiv:1603.02754). DOI: <https://doi.org/10.1145/2939672.2939785>.
12. Brownlee J. XGBoost for regression. *Machine Learning Mastery*. URL: <https://machinelearningmastery.com/xgboost-for-regression/>.
13. Kavlakoglu E., Russi E. What is XGBoost?. *IBM - United States*. URL: <https://www.ibm.com/think/topics/xgboost>.
14. Kaya H., Tufekci P. Gas turbine CO and NOx emission data set. *UCI Machine Learning Repository*. DOI: <https://doi.org/10.24432/C5WC95>.
15. StandardScaler. *scikit-learn*. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>.
16. Turney S. Coefficient of Determination (R^2) | Calculation & Interpretation. *Scribbr*. URL: <https://www.scribbr.com/statistics/coefficient-of-determination/>.
- Eng 2024), Targu Mures, Romania, Oct. 3–4, 2024. Cham: Springer, 2025, pp. 1–11, doi: https://doi.org/10.1007/978-3-031-81685-7_1
4. T. T. T. Han and C.-Y. Lin, “Exploring long-run CO₂ emission patterns and the environmental kuznets curve with machine learning methods”, *Innovation and Green Development*, vol. 4, no. 1, Feb. 2025, Art. no. 100195, doi: <https://doi.org/10.1016/j.igd.2024.100195>
5. Y. Hou and S. Liu, “Predictive modeling and validation of carbon emissions from China’s coastal construction industry: A BO-XGBoost ensemble approach”, *Sustainability*, vol. 16, no. 10, May 2024, Art. no. 4215, doi: <https://doi.org/10.3390/su16104215>
6. N. Gan and S. Zhao, “Global greenhouse gas reduction forecasting via machine learning model in the scenario of energy transition”, *Journal of Environmental Management*, vol. 371, Dec. 2024, Art. no. 123309, doi: <https://doi.org/10.1016/j.jenvman.2024.123309>
7. Y. Ma et al., “Machine learning-based prediction of the CO₂ concentration in the flue gas and carbon emissions from a waste incineration plant”, *ACS ES&T Engineering*, vol. 4, no. 3, pp. 737–747, Jan. 2024, doi: <https://doi.org/10.1021/acsestengg.3c00461>
8. G. Biau and E. Scornet, *A Random Forest Guided Tour*. To be published, doi: <https://doi.org/10.48550/arXiv.1511.05741>
9. “Random forest algorithm in machine learning.” *GeeksforGeeks*. [Online]. Available: <https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/>
10. J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, “A comparison of random forest variable selection methods for classification prediction modeling”, *Expert Systems With Applications*, vol. 134, pp. 93–101, Nov. 2019, doi: <https://doi.org/10.1016/j.eswa.2019.05.028>
11. T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*. To be published, doi: <https://doi.org/10.1145/2939672.2939785>
12. J. Brownlee. “XGBoost for regression.” *Machine Learning Mastery*. [Online]. Available: <https://machinelearningmastery.com/xgboost-for-regression/>
13. E. Kavlakoglu and E. Russi. “What is XGBoost?” *IBM - United States*. [Online]. Available: <https://www.ibm.com/think/topics/xgboost>
14. H. Kaya and P. Tufekci. “Gas turbine CO and nox emission data set.” *UCI Machine Learning Repository*, doi: <https://doi.org/10.24432/C5WC95>
15. “StandardScaler.” *scikit-learn*. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>
16. S. Turney. “Coefficient of Determination (R^2) | Calculation & Interpretation.” *Scribbr*. [Online]. Available: <https://www.scribbr.com/statistics/coefficient-of-determination/>

References

1. Committee of the Verkhovna Rada of Ukraine on Environmental Policy and Nature Management, “Informatsiino-analitychna zapyska shchodo stanu zakonodavstva z pytan zminy klimatu ta osnovnykh napriamkiv diialnosti Verkhovnoi Rady Ukrainy u tsii sferi [Information and analytical note on the state of legislation on climate change and the main areas of activity of the Verkhovna Rada of Ukraine in this field]”, To the letter of the Institute of Legislation of the Verkhovna Rada of Ukraine dated 27 April 2020 No. 1-17, 22/1612020. [Online]. Available: <https://komekolog.rada.gov.ua/uploads/documents/35959.pdf> (in Ukrainian)
2. “What is a gas turbine power station?” *Allied Power Group*. [Online]. Available: <https://alliedpg.com/latest-articles/gas-turbine-power-station/>
3. M. M. Muraru, Z. Simó, and L. B. Iantovics, “Study on state-of-the-art machine learning algorithms efficiency for CO₂ emissions prediction in fleet management optimization”, in *The 18th International Conference Interdisciplinarity in Engineering (Inter-*

Надійшла (received) 14.06.2025

UDC 621.438:519.254

DOROSHENKO DMYTRO ✉ – Master Student of the Department of Electric Power Transmission, National Technical University “Kharkiv Polytechnic Institute”; Kharkiv, Ukraine; e-mail: ddoroshenko55@gmail.com.

FORECASTING CARBON DIOXIDE (CO₂) EMISSIONS FROM GAS TURBINE PLANTS USING MACHINE LEARNING MODELS

This study explores the application of machine learning methods, specifically XGBoost and Random Forest models, for predicting carbon dioxide (CO₂) emissions from gas turbine equipment. Accurate forecasting of harmful emissions is a critical task in the context of ensuring energy security and achieving environmental sustainability. The research utilizes a real-world dataset collected in Turkey between 2011 and 2015, containing over 36,000 hourly monitoring records of technical and atmospheric parameters from a gas turbine installation. To enhance model performance, data preprocessing was conducted, including cleaning, feature correlation analysis, and standardization of numerical values. Based on the selected parameters, two regression models were developed: XGBRegressor and RandomForestRegressor, both demonstrating high accuracy in evaluation. The highest coefficient of determination (R^2) achieved was 0.742 for the Random Forest model, indicating its effectiveness in capturing the relationships between technical parameters and CO₂ emission levels. Feature importance analysis revealed that turbine temperature-related parameters had the greatest influence on emission levels, whereas external atmospheric conditions played a secondary role. The obtained results are practically significant for the energy sector, as they can be used for emission monitoring, planning the modernization of energy facilities, informing environmental policy, and selecting optimal operational modes for equipment. Implementing such models in industrial practice will contribute to reducing greenhouse gas emissions, increasing energy efficiency, and supporting the realization of national sustainable development strategies. The study confirms that the use of modern machine learning-based analytical tools can make a substantial contribution to ensuring a country's environmental and energy security.

Keywords: CO₂ emission forecasting; energy security; machine learning; Random Forest; XGBoost; regression model; gas turbine installation; data standardization.