



Кліматична нейтральність економіки та зелений перехід

Олена ЖИТКЕВИЧ,
Андрій МАТВІЙЧУК

**РОЗРОБЛЕННЯ ГІБРИДНОГО МЕТОДОЛОГІЧНОГО
ПІДХОДУ ДО ПРОГНОЗУВАННЯ
ДЕКАРБОНІЗАЦІЙНОГО ПОТЕНЦІАЛУ КРАЇНИ**

Резюме

Висока варіативність даних щодо декарбонізаційного потенціалу та недостатня формалізація цих процесів зумовлюють необхідність застосування сучасних гібридних методів аналізу. У дослідженні запропоновано підхід, що поєднує самоорганізаційні карти Кохонена (SOM) для виявлення закономірностей між країнами, а також лінійну регресію (LR) і нейронну мережу багат шарового перцептрона (MLP) для прогнозування середнього коефіцієнта викидів CO₂ у межах кластерів. Аналіз охоплює 45 країн за десятирічний період і ґрунтується на 12 пояснювальних показниках. Запропонована наукова методологія базується на інтеграції кластеризації та нейромережевого прогнозування для моделювання відмінностей у національних процесах декарбонізації. Карти Кохонена дали змогу виокремити сім груп країн, що підтверджує неоднорідність впливу факторів і закономірностей декарбонізації між країнами. Результати дослідження показали, що лінійні моделі мають обмежену пояснювальну здатність ($R^2 = 0,30$ на тестових даних до кластеризації та в середньому 0,61 після кластеризації), тоді як оптимізована нейронна мережа забезпечує вищу точність прогнозування (середнє R^2 для всіх кластерів становить 0,75 на тестових даних), що підтверджує доцільність кластерного моделювання та здатність нейронних мереж відображати нелінійні залежності. Отримані результати можуть бути використані розробниками політики для обґрунтування національної кліматичної політики та планування енергетичних переходів.

© Олена Житкевич, Андрій Матвійчук, 2026.

Житкевич Олена, кандидат економічних наук, доцент, Коледж бізнесу, Американський університет Курдистану, Духок, Ірак. ORCID: 0000-0003-2042-8795 Email: olena.zhytkevych@auk.edu.krd
Матвійчук Андрій, доктор економічних наук, професор, кафедра штучного інтелекту, моделювання та статистики, Київський національний економічний університет імені Вадима Гетьмана, Київ, Україна. ORCID: 0000-0002-8911-5677 Email: matviychuk@kneu.edu.ua

Ключові слова:

Data Science, викиди CO₂, зелена трансформація, карта Кохонена, нейронні мережі, нелінійне моделювання.

Класифікація за JEL: Q50, C12, C38, C53, C 55.

2 формули, 2 таблиці, 4 рисунки, 8 додатків, 36 джерел літератури.

Постановка проблеми

Для європейської економіки ключовими викликами є реалізація політики кліматичної нейтральності, перехід до низьковуглецевої моделі розвитку та узгодження національних кліматичних стратегій із довгостроковими цілями декарбонізації.

Ефективна реалізація низьковуглецевої політики залежить від стратегічного прогнозування декарбонізаційного потенціалу, що охоплює відбір, оцінювання та інтерпретацію екологічних, енергетичних, соціально-економічних та інших даних для обґрунтування управлінських рішень. Водночас європейські країни істотно відрізняються за структурою енергоспоживання, рівнем економічного розвитку, енергоємністю, електрифікацією тощо. Така структурна неоднорідність зумовлює відмінності в поточних рівнях вуглецевої інтенсивності та потенційних траєкторіях декарбонізації. Тому застосування єдиної моделі прогнозування для всіх країн може недостатньо враховувати специфічні закономірності окремих груп країн. Отже, подальший розвиток теоретичних, методологічних і практичних засад прогнозування декарбонізації є критично важливим для належного розвитку зеленої економіки.

Огляд літератури

Точне прогнозування рівня декарбонізації на національному рівні є важливим як для окремих країн, так і з глобальної перспективи. Водночас це складне завдання через нелінійну поведінку часових рядів середніх рівнів викидів та вплив непередбачуваних подій. Сучасні дослідження використовують широкий спектр методів, зокрема статистичні моделі, підходи просторово-часової візуалізації, моделі стаціонарних часових рядів (Encalada-Malca et al., 2021; Solci et al., 2023), алгоритми машинного навчання (Zhu et al., 2018), штучні нейронні мережі (Feng et al., 2015; Cordova et al., 2021; Fu et al., 2015; Kourentzes, et al., 2014; Wu et al., 2011; Paliwal & Kumar 2009), методи глибокого навчання (Li et al., 2020) та інші гібридні підходи.

Багатошарові перцептрони дають змогу моделювати складні залежності та застосовувалися для оцінювання впливу відновлюваних джерел енергії на викиди CO₂ (Mussa & Khalifa, 2025), однак вони є ресурсоемними та схильними

до перенавчання (Haykin, 2009). Для врахування довгострокових залежностей використовують LSTM, зокрема для прогнозування викидів CO₂ у Китаї та інших країнах (Wen et al., 2023; Kumari & Singh, 2023); CNN–LSTM застосовувалися до складних часових рядів в екологічних системах (Ding et al., 2021; Han et al., 2023). Водночас такі моделі потребують значно більших обчислювальних ресурсів порівняно з MLP.

Zheng et al. (2013) здійснювали оцінювання оперативної та деталізованої інформації про якість повітря в межах міста на основі історичних і поточних даних станцій моніторингу та інших міських джерел даних із використанням штучної нейронної мережі. Такі методи, однак, потребують значних обчислювальних ресурсів і великих масивів даних. Методи ансамблевого навчання, зокрема Random Forest, використовуються для прогнозування викидів парникових газів у сценаріях енергетичного переходу (Singh et al., 2026), але також є ресурсоємними (Breiman, 2001; Bielinskyi et al., 2026). Гібридні моделі, наприклад ARIMA у поєднанні з MLP або XGBoost, дають змогу одночасно моделювати тренди та нелінійності й застосовувалися для прогнозування викидів CO₂ у мегаполісах Китаю та країнах із перехідною економікою (Zhang et al., 2024; Zhang, 2003), проте потребують складної побудови й оптимізації. Метод Prophet продемонстрував ефективність для прогнозування нелінійних залежностей у моделях міських викидів та енергетичних сценаріях (Hasnain et al., 2022), однак є чутливим до налаштування гіперпараметрів (Taylor & Letham, 2017).

Проведений огляд літератури свідчить, що методи прогнозування декарбонізації мають різну придатність і обмеження: високі обчислювальні витрати, складність налаштування та ризик перенавчання за наявності коротких часових рядів для окремих країн. Гібридні моделі, що поєднують статистичний аналіз і машинне навчання, є перспективними для країн із трансформаційними економіками та перехідним енергетичним ландшафтом. Точність і стабільність результатів відрізняються залежно від масштабу та типу даних. Ресурсоємність, потреба у великих масивах даних і низька інтерпретованість ускладнюють адаптацію моделей, що зумовлює необхідність розроблення більш універсальних та ефективних підходів до прогнозування декарбонізаційного потенціалу.

Метою статті є розроблення та емпірична апробація гібридного підходу до прогнозування декарбонізаційного потенціалу, який передбачає поєднане застосування кластеризації країн (карт Кохонена) та внутрішньокластерного моделювання з використанням багатозарового перцептрона і лінійної регресії.

Для досягнення поставленої мети необхідно розв'язати такі завдання:

1. Попередня обробка даних і перевірка гіпотез щодо ключових показників – здійснити попередню обробку даних (логарифмування цільової змінної, аналіз мультиколінеарності) та перевірити гіпотези щодо значущості ключових факторів;
2. Розроблення гібридного підходу, що інтегрує кластеризацію та нелінійне моделювання динамічних показників: (2.1) кластеризація країн за структурними характеристиками із застосуванням SOM; (2.2) порівняльне оцінювання регресійної та MLP-моделей з урахуванням кластерної структури.

Методологія

У дослідженні застосовано гібридну кількісну методологію, що поєднує попередню обробку даних, статистичну діагностику, кластеризацію та прогнозне моделювання. Панельний набір даних для 45 країн за 2014–2023 рр. містить 12 пояснювальних показників і середній коефіцієнт викидів CO₂ як цільову змінну. Дані було очищено, нормалізовано до інтервалу [0–1], а цільову змінну логарифмічно трансформовано. Мультиколінеарність оцінювали за допомогою кореляції Пірсона та VIF. Профілі країн кластеризували із застосуванням самоорганізаційних карт Кохонена. Далі для повної вибірки та в межах кластерів оцінювали моделі лінійної регресії й багатошарового перцептрона з поділом даних на навчальну і тестову вибірки у співвідношенні 70/30 та оцінюванням за R², MAE і MSE.

Результати дослідження

Завдання 1. Попередня обробка даних і перевірка гіпотез щодо ключових показників.

Прогнозування є формою наукового передбачення, що має два аспекти: описовий, який оцінює можливі майбутні сценарії, та прескриптивний (наказовий), який підтримує управлінські рішення. Воно зменшує невизначеність і дає змогу визначати наслідки дій або бездіяльності, забезпечуючи причинно-наслідковий зв'язок між поточним станом і перспективами розвитку. Ефективна політика декарбонізації країни потребує інструментів для оцінювання поточного потенціалу та прогнозування його динаміки з урахуванням внутрішніх і зовнішніх факторів.

У дослідженні аналізується декарбонізаційний потенціал 45 країн за період 2014–2023 рр. Масив даних містить 13 річних показників із соціально-економічної, енергетичної, екологічної, глобалізаційної та геополітичної сфер, які сукупно відображають декарбонізаційний потенціал країни. Показники відібрано з EnerData (2026), World Bank (n. d.) та Geopolitical Risk Index (n. d.) на основі аналізу мультиколінеарності й значущості в попередніх дослідженнях (Zhytkevych et al., 2022, 2023). Вибірка охоплює країни OECD, G7, BRICS, EU, CIS, а також регіони Північної та Латинської Америки, Азії, Австралії, Африки й Близького Сходу. Обмеження вибірки 45 країнами забезпечує глобальну репрезентативність і включає країни з різним рівнем залученості до процесів декарбонізації, для яких доступний достатній обсяг релевантних статистичних даних.

Дослідження Vanpa et al. (2023) показує, що геополітична напруженість, зокрема вторгнення Росії в Україну, дестабілізує енергетичні системи, фінансові ринки та інвестиції в зелені технології, зміщуючи пріоритети у бік безпеки й оборони. Геополітичні ризики є важливою змінною для оцінювання глобальної декарбонізації. Індекс геополітичного ризику (GRI), описаний Caldara & Iacoviello (2022) і побудований на аналізі медійного висвітлення конфліктів, відображає їхній вплив на глобальний геополітичний ландшафт та економічну динаміку. У цьому дослідженні GRI розглядається як один із факторів прогнозування де-

карбонізаційного потенціалу, оскільки він впливає на енергетичну систему та спроможність країни скорочувати викиди CO_2 . Для Узбекистану, де був відсутній один показник (GRI), його значення замінено середнім значенням для регіону Центральної Азії (Казахстан, Киргизстан, Таджикистан) за 10 років, що дає змогу зберегти повноту даних.

$$D = \{X^k = x_{i,j}^k \mid k = 1, \dots, K; i = 1, \dots, n; j = 1, \dots, m\}, \quad (1)$$

де $k = 1, \dots, K$ – індекс країни ($K = 45$); $i = 1, \dots, n$ – часовий індекс річних спостережень ($n = 10$); $j = 1, \dots, m$ – індекс показника ($m = 13$).

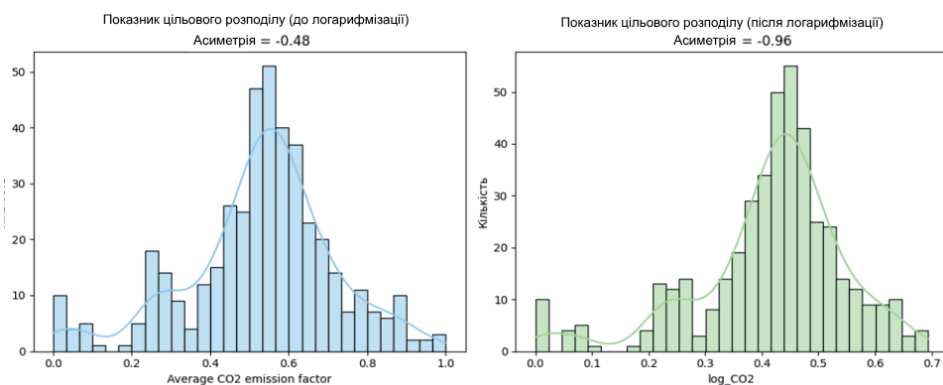
У всіх подальших експериментах із моделювання набір даних було поділено на навчальну та тестову (валідаційну) підвибірki у співвідношенні 70/30.

Серед 13 відібраних показників цільовою змінною є середній коефіцієнт викидів CO_2 . Декарбонізаційний потенціал країни інтерпретується як її оцінена спроможність скорочувати викиди CO_2 , що формується під впливом рівня технологічного розвитку, потужностей відновлюваної енергетики, енергоефективності, а також економічного й демографічного зростання. Отже, цей показник можна розглядати як інтегральний індикатор, що відображає сукупний вплив інших детермінант. Решта 12 показників інтерпретуються як пояснювальні змінні, які з різною інтенсивністю пов'язані з декарбонізаційним потенціалом. Припускається, що частина факторів може мати неоднозначний вплив, тоді як для більшості очікується визначений напрям зв'язку. У сукупності вони характеризують ключові напрями формування декарбонізаційного потенціалу країни та є основою для формулювання дослідницьких гіпотез (див. Додаток А).

Позначимо декарбонізаційний потенціал y_i^k країни k у році i , де цільовою змінною y_i^k є середній коефіцієнт викидів CO_2 $x_{i,12}^k$ а $X_i^k = (x_{i,1}^k, \dots, x_{i,12}^k)$ – вектор пояснювальних змінних містить 12 показників, наведених у Додатку А. Загальну нульову гіпотезу сформульовано так: $H_0: \beta_1 = \beta_2 = \dots = \beta_{12} = 0$ – жоден із показників не має статистично значущого впливу на цільову змінну y_i^k . Альтернативна гіпотеза $H_1: \exists j \in \{1, \dots, 12\}, \beta_j \neq 0$ передбачає, що принаймні один із показників має статистично значущий вплив на y_i^k .

Після підготовки даних (очищення, форматування, заповнення пропусків, нормалізація до $[0-1]$) та визначення цільового показника розраховано описову статистику для 13 показників, наведену в Додатку В. Середнє значення коефіцієнта викидів CO_2 ($x_{1,12}$) становить 0,531 і є близьким до медіани; стандартне відхилення 0,186 свідчить про відмінності в інтенсивності викидів між країнами. Аналіз розподілу та асиметрії виявив необхідність трансформації цільової змінної: CO_2 було логарифмовано з додаванням константи для уникнення нульових значень, що звузило розподіл і зменшило вплив екстремальних значень. Трансформований показник позначено як $\log_ \text{CO}_2$ (рис. 1).

Рисунок 1

Розподіл середнього коефіцієнта викидів CO₂ до та після логарифмування

Джерело: розраховано авторами.

Для оцінювання значущості та напрямку впливу відібраних пояснювальних змінних застосовано модель множинної лінійної регресії. Спочатку було специфіковано теоретичну регресійну модель загального вигляду $\log_CO_2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{12} x_{12} + \varepsilon$, де β – коефіцієнти регресії, а ε – випадкова похибка.

Множинну лінійну регресію реалізовано на навчальному наборі даних у Power BI за допомогою Python-скрипту із застосуванням методу Linear Regression() пакета scikit-learn. Параметри моделі оцінено методом найменших квадратів, який мінімізує суму квадратів відхилень прогнозних значень від фактичних. У результаті отримано оцінене рівняння регресії:

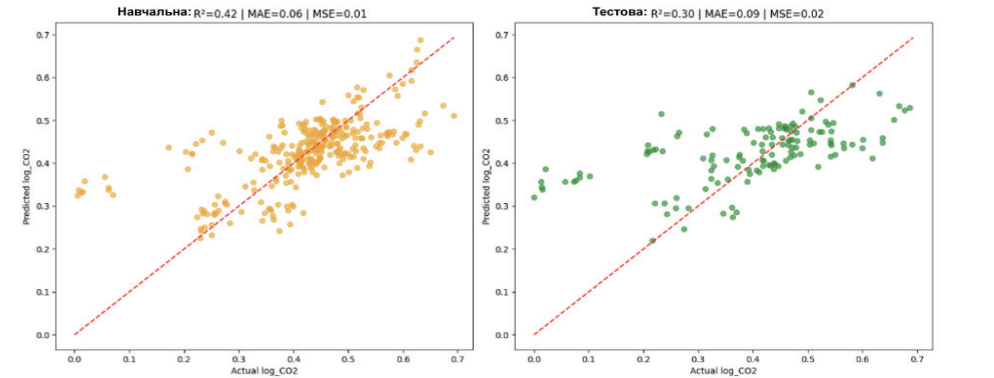
$$\hat{y} = 0.258 - 0.019x_1 + 0.048x_2 - 0.846x_3 - 0.305x_4 + 0.137x_5 - 0.002x_6 + 0.149x_7 - 0.786x_8 + 0.200x_9 - 0.051x_{10} + 0.162x_{11} - 0.020x_{12} \quad (2)$$

Перелік показників наведено в табл. 1. Рівняння (2) кількісно визначає граничний вплив кожної пояснювальної змінної на логарифм середнього коефіцієнта викидів CO₂ за незмінності інших 12 показників.

Якість моделі (рис. 2) оцінено на навчальній вибірці: $R^2 = 0,418$, MAE = 0,060, MSE = 0,008; на тестовій вибірці: $R^2 = 0,299$, MAE = 0,090, MSE = 0,016, що свідчить про помірну пояснювальну здатність на навчальній вибірці та ще нижчу – на тестовій. Загальну значущість моделі підтверджено F-тестом ($p \approx 3,6 \times 10^{-31} < 0,001$), що дає підстави відхилити нульову гіпотезу про рівність усіх коефіцієнтів регресії нулю. Перевірку гіпотез за знаком β (див. Додаток А) та рівнем значущості ($p\text{-value} < 0,1$) проведено на повному наборі даних для опису поточних емпіричних зв'язків; модель оцінено за всіма спостереженнями для максимального використання інформації та отримання точніших оцінок коефіцієнтів (табл. 1).

Рисунок 2

Співвідношення фактичних і прогнозних значень $\log_CO_2$ у навчальній та тестовій вибірках за моделлю лінійної регресії



Джерело: розраховано авторами.

Таблиця 1

Перевірка гіпотез на повному наборі даних

Показник	β	p-value	Очікуваний знак (H_1)	Висновки
Енергоємність ВВП за споживанням	0,03	0,47062	>0	Незначущий, H_1 відхилено
Внутрішнє споживання нафтопродуктів	0,19	0,34053	>0	Незначущий, H_1 відхилено
Внутрішнє споживання природного газу	0,42	0,00192	>0	Значущий, H_1 прийнято
Внутрішнє споживання електроенергії	-1,56	0,00038	Будь-який	Значущий, H_1 прийнято
Частка відновлюваних джерел в електрифікації	-0,49	1×10^{-5}	<0	Значущий, H_1 прийнято
Частка вітрової та сонячної енергії в електрифікації	0,30	1×10^{-5}	<0	Значущий, але знак протилежний очікуваному
Зростання міського населення (річне, %)	0,05	0,55406	Будь-який	Незначущий, H_1 відхилено
Зростання ВВП на душу населення (річне, %)	0,14	0,06451	Будь-який	Значущий, H_1 прийнято
Внутрішнє споживання вугілля та лігніту	1,36	1×10^{-5}	>0	Значущий, H_1 прийнято
Електрифікація	0,36	1×10^{-5}	Будь-який	Значущий, H_1 прийнято
Індекс геополітичних ризиків	-0,06	0,29281	Будь-який	Незначущий, H_1 відхилено
Індекс глобалізації	0,28	0,00008	Будь-який	Значущий, H_1 прийнято

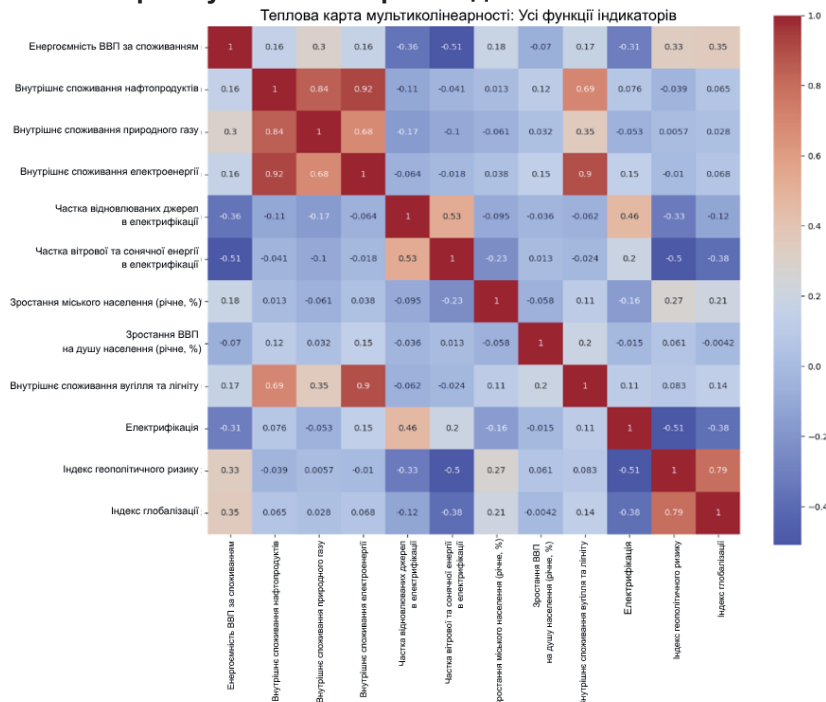
Джерело: розраховано авторами.

У табл. 1 наведено результати регресійного аналізу: сім показників є значущими та підтверджують H1 (виділено напівжирним), чотири показники статистично незначущі ($p > 0,1$), навіть якщо знак β відповідає гіпотезі, а один показник (частка вітрової та сонячної енергії в електрифікації) є значущим, проте знак суперечить гіпотезі. Загальна перевірка гіпотези є статистично значущою; H0 відхилено, отже принаймні один із 12 показників має статистично значущий зв'язок із $\log_CO_2$.

Хоча загальна модель має певну пояснювальну здатність, статистична незначущість окремих предикторів може бути пов'язана з мультиколінеарністю або проблемами специфікації, що потребують трансформації змінних. У дослідженні значущі показники відбиралися на основі поєднання кореляційного аналізу та фільтрації за фактором інфляції дисперсії (VIF – Variance Inflation Factor), відповідно до сучасних рекомендацій щодо запобігання мультиколінеарності в моделях прогнозування викидів та енергетичного моделювання (Fuinhas et al., 2025). Для оцінювання мультиколінеарності між 12 показниками декарбонізації побудовано кореляційну теплову карту та розраховано VIF (Belsley et al., 1980). Кореляційну теплову карту побудовано в Microsoft Power BI за допомогою скрипту Python 3.10 із бібліотеками pandas, matplotlib і seaborn. Вона відображає коефіцієнти кореляції Пірсона між усіма парами показників (рис. 3).

Рисунок 3

Теплова карта мультиколінеарності для 12 показників



Джерело: розраховано авторами.

На основі розрахованих коефіцієнтів кореляції для 12 показників (рис. 3) виявлено змінні з високою мультиколінеарністю. Далі в Microsoft Power BI за допомогою Python (statsmodels) розраховано VIF, який характеризує завищення дисперсії коефіцієнтів унаслідок колінеарності ($VIF < 5$ – низька; від 5 до 10 – помірна; понад 10 – висока мультиколінеарність). Початкові значення VIF для внутрішнього споживання електроенергії (149,62) та зростання ВВП на душу населення (114,61) свідчили про сильну колінеарність.

Відбір ознак здійснювали шляхом поетапного вилучення змінних із найвищими значеннями VIF, доки для всіх інших показників VIF не став ≤ 10 , що зменшило надлишковість і зберегло інформаційне різноманіття. У результаті було виключено 9 показників: внутрішнє споживання електроенергії, зростання ВВП на душу населення, зростання міського населення, внутрішнє споживання нафтопродуктів, індекс глобалізації, внутрішнє споживання вугілля та лігніту, індекс геополітичного ризику, електрифікацію та внутрішнє споживання природного газу.

Вилучення змінних тільки на підставі статистичної мультиколінеарності потребує експертного контролю, оскільки реальні взаємозв'язки між факторами декарбонізації можуть бути нелінійними або залежними від контексту.

Завдання 2. Розроблення гібридного підходу, що інтегрує кластеризацію та нелінійне моделювання динамічних показників.

2.1. Кластеризація країн за структурними характеристиками із застосуванням SOM.

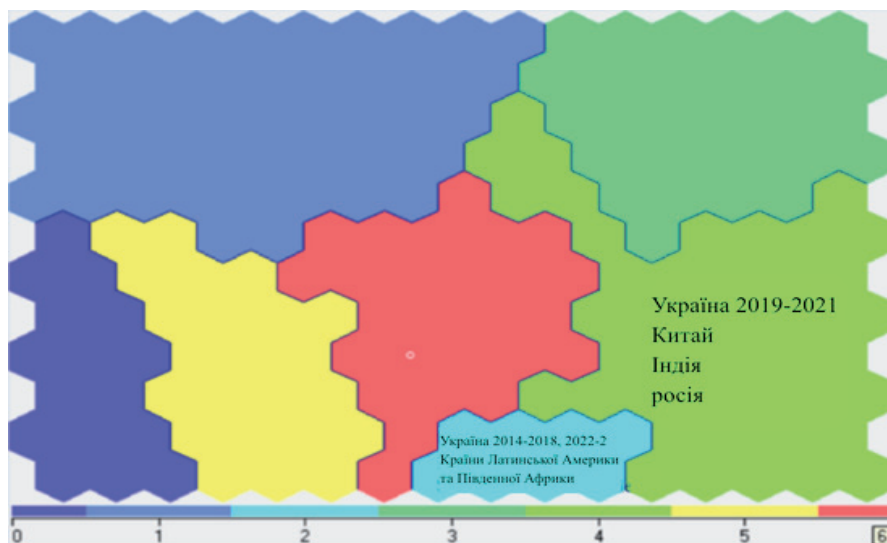
Наступним етапом дослідження є аналіз нелінійних взаємозв'язків між показниками, які неможливо повною мірою визначити за допомогою лінійних діагностичних методів. Важливо враховувати істотні відмінності в енергетичному та економічному розвитку різних країн. Наприклад, відмінності в динаміці розвитку Китаю, Нідерландів та Узбекистану є очевидними. Тому доцільно виокремити групи подібних країн і будувати прогностні моделі в межах цих груп. Відповідно, країни із загальної вибірки кластеризуються, для кожного кластера будуються прогностні моделі (лінійна регресія та MLP), а також єдині моделі LR і MLP для всіх країн; їхню результативність порівнюють для обґрунтування доцільності запропонованого підходу.

Серія експериментів показала, що оптимально використовувати 12 показників. Кластеризація на основі всіх 12 змінних забезпечила стабільніші кластери, тому повний набір показників збережено в подальшому аналізі. Це також узгоджується з властивостями SOM, які ефективно працюють із багатовимірними та нелінійно пов'язаними даними, зберігаючи топологічну й метричну структуру вхідного простору (Peeters & Dassargues, 2006; Milone et al., 2013; Agrafiotis & Xu, 2002).

Порівняння лінійної та MLP-моделей на рівні кластерів має кілька переваг. По-перше, воно дає змогу перевірити, чи враховує нелінійна MLP-модель додаткову інформацію, яку не здатна відобразити лінійна модель, особливо в групах країн із нестандартними або екстремальними профілями показників. По-друге, аналіз у межах кластерів допомагає виявити локальну неоднорідність даних та оцінити адаптивність моделей до різних контекстів. Нарешті, кластери-

Отже, запропонований підхід забезпечує об'єктивний, ефективний і прозорий процес підтримки рішень у сфері декарбонізації. У дослідженні країни кластеризовано за набором із 12 показників декарбонізаційного потенціалу (перелік показників наведено в Додатку А; результати кластеризації – у Додатку С), а методологія кластеризації ґрунтується на підходах, розроблених і апробованих у праці Zhytkevych & Brochado (2022). Базу даних для аналізу оновлено, переглянуто й розширено шляхом включення показників ризику та глобалізації, а також збільшення вибірки до 45 країн, що підвищило репрезентативність оцінювання національного потенціалу скорочення викидів CO₂.

Самоорганізаційна карта із семи кластерів для 45 країн на основі 12 показників



Джерело: розраховано авторами.

Результати кластеризації засвідчили значну неоднорідність європейських країн за енергетичними, економічними та іншими характеристиками, що визначають декарбонізаційний потенціал. Виявлені кластери характеризуються як стабільним членством окремих країн, так і переходами між групами протягом 2014–2023 рр. Наприклад, кластер 0 характеризується стабільним перебуванням Норвегії та Швеції протягом усього десятиріччя. Кластер 1 включає Францію та Чехію у 2014–2023 рр., тоді як Бельгія, Велика Британія, Нідерланди, Німеччина та Польща належали до нього в окремі періоди. Кластер 5 об'єднує низку

європейських країн, зокрема Данію та Португалію протягом усього періоду, а також Бельгію, Велику Британію, Іспанію, Італію, Нідерланди та Німеччину в окремі роки. Кластер 6 включає Румунію у 2014–2023 рр., а також Іспанію, Італію, Німеччину та Польщу в окремі періоди (Додаток С).

Можливість спостерігати переходи країн між кластерами за допомогою SOM ілюструє чутливість моделі до змін енергетичних структур та економічного розвитку. Наприклад, Україна перебувала у кластері 2 у 2014–2018 рр. разом із країнами Латинської Америки та Південної Африки, у 2019–2021 рр. перейшла до кластера 4, де розміщувалися Індія, Китай та росія, а у 2022–2023 рр. повернулася до кластера 2. Подібні переходи спостерігалися для Індії, Мексики та інших країн, відображаючи зміни їхніх енергетичних профілів. Таким чином, SOM допомагає враховувати як загальні тенденції, так і специфічні для окремих країн зміни.

2.2. Порівняльне оцінювання регресійної та MLP-моделей з урахуванням кластерної структури

Після кластеризації дослідження переходить до побудови прогнозних моделей на двох рівнях: для повного набору даних і окремо для кожного виявленого кластера. Кластер-специфічні моделі оцінюються на основі спостережень країн, що належать до відповідного кластера протягом аналізованого періоду, що дає змогу оцінити доцільність кластеризації як інструменту виявлення неоднорідності між країнами.

Нейронну мережу реалізовано, оцінено та візуалізовано в Power BI за допомогою скрипту Python 3.10 із бібліотеками scikit-learn, numpy, pandas і matplotlib. Набір пояснювальних факторів складався з 12 показників, тоді як середній коефіцієнт викидів CO₂ використовувався як залежна змінна (Додаток А); для кожного виявленого кластера дані випадково поділялися на навчальну та тестову підвибірki.

Оптимізацію параметрів мережі здійснювали шляхом експериментів із систематичною зміною архітектури (кількості прихованих шарів і нейронів у кожному шарі) та кількості ітерацій навчання. Порівняльні експерименти показали, що найстабільнішу результативність за R², MAE та MSE забезпечила архітектура з чотирма шарами. Фрагмент Python-скрипту, що реалізує цю функцію, наведено в Додатку D. Зокрема, архітектура [12, 5, 6, 1] забезпечила найбільш стабільну апроксимацію нелінійних залежностей і збіжність моделі на навчальній і тестовій вибірках повного набору даних.

Для порівняння лінійна модель продемонструвала нижчі результати, ніж MLP (табл. 2). Це підтверджує, що традиційні лінійні методи не повною мірою відображають складність досліджуваних процесів. Незважаючи на окремі обмеження, показники MLP виявилися кращими, що підкреслює переваги нейронних мереж у моделюванні складних залежностей.

Таблиця 2

Порівняння результативності моделей LR та MLP на повному наборі даних

Модель	Навчальні дані			Тестові дані		
	R ²	MAE	MSE	R ²	MAE	MSE
LR (12 індикаторів)	0,418	0,060	0,008	0,299	0,090	0,016
MLP (12 індикаторів)	0,851	0,041	0,003	0,744	0,043	0,003

Джерело: розраховано авторами.

Результати серії проведених експериментів показали, що якість моделі істотно залежить не лише від кількості показників, а й від її архітектури. Експериментальні дослідження дали змогу точніше оцінити вплив конфігурацій на ефективність моделі та визначити оптимальну архітектуру MLP для 12 показників на повному наборі даних.

Наступним кроком стала побудова окремих MLP- і лінійних моделей для кожного кластера країн, що дає змогу підвищити точність прогнозування та врахувати специфіку внутрішньокластерних залежностей. Для семи кластерів, отриманих за допомогою SOM, моделі LR і MLP розраховано на основі Python-коду з використанням бібліотек `scikit-learn`, `numpy`, `pandas` і `matplotlib`. Для кожного кластера здійснено пошук оптимальної архітектури прогнозувальної нейронної мережі: [12, 6, 4, 1] для кластерів 0, 4 і 5; [12, 9, 6, 1] для кластерів 1 і 3; [12, 8, 6, 1] для кластера 2; [12, 4, 3, 1] для кластера 6. Оцінювання точності прогнозування для семи кластерів на основі LR та MLP з оптимальною архітектурою для кожного кластера графічно представлено в Додатку Е та узагальнено в табл. Е1.

Останній стовпець табл. Е1 містить середні значення показників результативності LR та MLP для всіх семи кластерів. Аналіз результатів регресійного та MLP-моделювання для кластерів 0–6 продемонстрував відмінності в точності прогнозування залежно від типу моделі та характеристик кластера. Загалом MLP-моделі показали вищі значення коефіцієнта детермінації R² та нижчі значення MAE і MSE порівняно з лінійними регресіями, особливо для тестових наборів даних, що свідчить про більшу здатність нейронних мереж відтворювати нелінійні взаємозв'язки між енергетичними та економічними показниками.

Це свідчить про те, що взаємозв'язки між показниками відрізняються між кластерами країн, що обґрунтовує використання кластерно-орієнтованих прогнозних моделей.

Після кластеризації MLP демонструє перевагу за результатами тестової вибірки для кластерів 2, 3, 4 і 6, підтверджуючи здатність ефективніше відображати складні нелінійні залежності порівняно з LR. Натомість у кластерах 0, 1 і 5 перевага MLP є обмеженою або відсутньою, що свідчить про достатність простіших лінійних моделей для цих груп, а також може вказувати на структурну однорідність або вищу варіативність, яка обмежує точність прогнозування. Значущі показники, визначені для кожного кластера, та відповідні рекомендовані методи моделювання узагальнено в Додатках F, G. Результати перевірки гіпотез для кожного кластера на основі оцінених зв'язків між пояснювальними показниками та цільовою змінною наведено в Додатку H.

Обговорення

Аналіз регресійних рівнянь, наведених у Додатку F, демонструє специфічний вплив окремих показників у кожному з 7 кластерів, що підтверджує наявність вираженої неоднорідності у структурі показників в межах загального набору даних. Таким чином, у таблиці в Додатку G наведено підсумок ключових індикаторів для кожного кластера та рекомендовані методи моделювання на основі порівняльного аналізу ефективності моделей і релевантності показників. Індикатори було відібрано на основі ефективності прогнозування (R^2 , MAE, MSE) та їхньої узгодженості в найкращих моделях кожного кластера. Цей аналіз підтверджує, що коефіцієнти викидів CO_2 суттєво різняться між кластерами – від традиційної залежності від викопного палива до впровадження відновлюваних джерел енергії, технологічно-інституційної трансформації та соціально-економічних чинників, що підкреслює необхідність розробки стратегій моделювання, специфічних для кожного кластера.

За прикладом перевірки гіпотез, описаних у Додатку A, для регресійної моделі, побудованої на повній вибірці даних (табл. 1), ми перевіримо ці гіпотези у регресійних моделях, побудованих для кожного з 7 кластерів. Схожим чином, під час перевірки гіпотез враховувалися як значення p , так і знак β . Гіпотеза вважалася підтвердженою лише в тому випадку, якщо оцінений ефект був статистично значущим ($p < 0,1$) і відповідав очікуваному знаку (Додаток A), про що свідчить виділення жирним шрифтом у Додатку H. Результати аналізу розрахунків, наведених у Додатку H, загалом підтвердили висновки, отримані для повного набору даних (табл. 1). Хоча з 12 гіпотез, висунутих у Додатку A щодо впливу досліджуваних факторів на залежну змінну, усі 12 були частково підтверджені принаймні в одному з кластерів, лише 7 гіпотез реалізувалися в більшості кластерів. Наприклад, неістотний показник x_1 «Енергоємність ВВП за споживанням» у моделі на повному наборі даних (табл. 1) виявився істотним лише для одного кластера з семи (кластер 6, що можна побачити в Додатку H), що також підтверджує його загальну неістотність.

Отримані результати підтверджують доцільність використання гібридного підходу, який поєднує кластеризацію країн за допомогою самоорганізаційних карт Кохонена та внутрішньокластерне моделювання. Наукова новизна підходу полягає в інтеграції кластеризації та прогнозного моделювання в єдину послідовну процедуру. SOM дає змогу виявляти відносно однорідні групи країн за сукупністю енергетичних, економічних і технологічних характеристик, тоді як використання MLP у сформованих кластерах дає змогу враховувати нелінійні залежності між показниками.

Дослідження має практичне значення для європейських країн, які характеризуються різними енергетичними структурами, рівнями економічного розвитку та темпами переходу до низьковуглецевої економіки.

Висновки

Декарбонізація є складним і неоднорідним процесом у різних країнах через відмінності в їхньому соціально-економічному, енергетичному та технологічному потенціалі. Наявність коротких часових рядів даних і швидкі зміни

енергоспоживання в окремих країнах істотно ускладнюють точне моделювання динаміки викидів CO_2 . Для подолання цих обмежень запропоновано гібридну модель, що поєднує кластеризацію з нелінійним прогнозуванням і забезпечує точніше оцінювання тенденцій для підтримки розроблення ефективних стратегій сталого розвитку. Сформовано набір даних для 45 країн за 2014–2023 рр. із 12 показників та середнього коефіцієнта викидів CO_2 як цільової змінної, що охоплює технологічні, енергетичні, економічні й демографічні фактори. Підготовка даних включала стандартизацію, заповнення пропусків і логарифмування цільової змінної для зменшення впливу екстремальних значень.

Аналіз мультиколінеарності та експертне оцінювання результатів кластеризації забезпечили збереження 12 показників. Множинна лінійна регресія продемонструвала помірну пояснювальну здатність ($R^2 = 0,42$ на навчальній і $0,30$ на тестовій вибірках), а F-тест підтвердив загальну значущість моделі. Для виявлення складних нелінійних залежностей використано багатосаровий перцептрон з оптимізованою архітектурою [12, 5, 6, 1], який забезпечив достатньо високу точність ($R^2 = 0,85$ на навчальній і $0,74$ на тестовій вибірках). Після кластеризації середні показники для семи кластерів додатково засвідчили поліпшення точності прогнозування: для LR середнє $R^2 = 0,61$, MAE = $0,06$, MSE = $0,01$; для MLP $R^2 = 0,75$, MAE = $0,03$, MSE = 1×10^{-4} на тестових даних.

Кластеризація країн на основі набору даних виявила сім кластерів із різними енергетичними, економічними та соціальними структурами. Вплив окремих показників відрізняється між кластерами, що обґрунтовує диференційований підхід до моделювання.

Результати показують, що всі 12 гіпотез були підтверджені лише частково для окремих кластерів; отже, не всі фактори мають стабільно значущий вплив, а напрям і сила зв'язків істотно відрізняються між кластерами. Це свідчить про неоднорідність взаємозв'язків між показниками та викидами CO_2 і підтримує застосування кластерного моделювання.

Практичне застосування отриманих результатів може слугувати основою для подальшого розроблення національних стратегій декарбонізації, визначення напрямів розвитку відновлюваної енергетики та енергоефективності, а також моніторингу прогресу країн у контексті глобальних цілей сталого розвитку.

Подальші дослідження передбачають використання альтернативних нейронних мереж та ансамблевого навчання для кластерно-орієнтованої інтерпретації результатів моделей і валідації гібридного прогнозного підходу, підвищення точності прогнозування, проведення аналізу чутливості моделей, регіонального прогнозування декарбонізації та інтеграції в системи підтримки рішень для енергетичної політики й кліматичних стратегій.

Подяка

Роботу виконано в межах прикладного наукового дослідження «Дата Сайнс аналіз зеленої трансформації глобальної продовольчої безпеки та імплементація міжнародного досвіду в Україні», що фінансується Міністерством освіти і науки України (проект № 0126U001738).

Список використаної літератури

- Agrafiotis, D. K., & Xu, H. (2002). A self-organizing principle for learning nonlinear manifolds. *Proceedings of the National Academy of Sciences*, 99(25), 15869–15872. <https://doi.org/10.1073/pnas.242424399>
- Banna, H., Alam, A., Chen, X. H., & Alam, A. W. (2023). Energy security and economic stability: The role of inflation and war. *Energy Economics*, 126, Article 106949. <https://doi.org/10.1016/j.eneco.2023.106949>
- Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. Wiley. <https://doi.org/10.1002/0471725153>
- Bielinskiy, A., Soloviev, V., Matviychuk, A., & Bezkorovainyi, V. (2026). Stock market movement forecasting using machine learning and complex network measures. *Artificial Intelligence and Applications*. <https://doi.org/10.47852/bonviewAIA62027537>
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Caldara, D., & Iacoviello, M. (2022). Measuring geopolitical risk. *American Economic Review*, 112(4), 1194–1225. <https://doi.org/10.1257/aer.20191823>
- Cordova, C. H., Portocarrero, M. N. L., Salas, R., Torres, R., Rodrigues, P. C., & López-Gonzales, J. L. (2021). Air quality assessment and pollution forecasting using artificial neural networks in Metropolitan Lima–Peru. *Scientific Reports*, 11, Article 24232. <https://doi.org/10.1038/s41598-021-03650-9>
- Ding, C., Wang, G., Zhang, X., Liu, Q., & Liu, X. (2021). A hybrid CNN-LSTM model for predicting PM_{2.5} in Beijing based on spatiotemporal correlation. *Environmental and Ecological Statistics*, 28, 503–522. <https://doi.org/10.1007/s10651-021-00501-8>
- Encalada-Malca, A. A., Cochachi-Bustamante, J. D., Rodrigues, P. C., Salas, R., & López-Gonzales, J. L. (2021). A spatio-temporal visualization approach of PM₁₀ concentration data in Metropolitan Lima. *Atmosphere*, 12(5), Article 609. <https://doi.org/10.3390/atmos12050609>
- EnerData. (2026). *Total energy consumption. World Energy & Climate Statistics – Yearbook 2026* [dataset]. <https://yearbook.enerdata.net/total-energy/world-consumption-statistics.html>
- Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., & Wang, J. (2015). Artificial neural networks forecasting of PM_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment*, 107, 118–128. <https://doi.org/10.1016/j.atmosenv.2015.02.030>
- Fu, M., Wang, W., Le, Z., & Khorram, M. S. (2015). Prediction of particular matter concentrations by developed feed-forward neural network with rolling mechanism and gray model. *Neural Computing and Applications*, 26, 1789–1797. <https://doi.org/10.1007/s00521-015-1853-8>
- Fuinhas, J. A., Belucio, M., Santiago, R., & Betencourt, M. (2025). Are the structure dynamics of capital stock impacting carbon intensity from energy

- consumption? European insights. *Energy, Ecology and Environment*, 10, 555–574. <https://doi.org/10.1007/s40974-025-00370-y>
- Geopolitical Risk Index. (n. d.). *Final Geopolitical Risk Index* [dataset]. <https://www.geopriskindex.com/results-final-risk-index/>
- Han, Z., Cui, B., Xu, L., Wang, J., & Guo, Z. (2023). Coupling LSTM and CNN neural networks for accurate carbon emission prediction in 30 Chinese provinces. *Sustainability*, 15(18), Article 13934. <https://doi.org/10.3390/su151813934>
- Hasnain, A., Sheng, Y., Hashmi, M. Z., Bhatti, U. A., Hussain, A., Hameed, M., Marjan, S., Bazai, S. U., Hossain, M. A., Sahabuddin, M., Wagan, R. A., & Zha, Y. (2022). Time series analysis and forecasting of air pollutants based on Prophet forecasting model in Jiangsu Province, China. *Frontiers in Environmental Science*, 10, Article 945628. <https://doi.org/10.3389/fenvs.2022.945628>
- Haykin, S. S. (2009). *Neural Networks and Learning Machines* (3rd ed.). Pearson.
- Kourentzes, N., Barrow, D. K., & Crone, S. F. (2014). Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, 41(9), 4235–4244. <https://doi.org/10.1016/j.eswa.2013.12.011>
- Kumari, S., & Singh, S. K. (2023). Machine learning–based time series models for effective CO₂ emission prediction in India. *Environmental Science and Pollution Research*, 30(55), 116601–116616. <https://doi.org/10.1007/s11356-022-21723-8>
- Li, S., Xie, G., Ren, J., Guo, L., Yang, Y., & Xu, X. (2020). Urban PM_{2.5} concentration prediction via attention-based CNN–LSTM. *Applied Sciences*, 10(6), Article 1953. <https://doi.org/10.3390/app10061953>
- Milone, D. H., Stegmayer, G., Kamenetzky, L., López, P., & Carrari, F. (2013). Clustering biological data with SOMs: On topology preservation in non-linear dimensional reduction. *Expert Systems with Applications*, 40(9), 3841–3845. <https://doi.org/10.1016/j.eswa.2012.12.074>
- Mussa, A. A. Y., & Khalifa, W. M. S. (2025). MLP enhanced CO₂ emission prediction model with LWSSA nature inspired optimization. *Scientific Reports*, 15, Article 1891. <https://doi.org/10.1038/s41598-025-85709-5>
- Paliwal, M., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, 36(1), 2–17. <https://doi.org/10.1016/j.eswa.2007.10.005>
- Peeters, L., & Dassargues, A. (2006). *Comparison of Kohonen's self-organizing map algorithm and principal component analysis in the exploratory data analysis of a groundwater quality dataset* (Research Report). University of Liège. <https://orbi.uliege.be/bitstream/2268/3388/1/publi144-2007.pdf>
- Singh, S., Kumar, M., Verma, B. K., & Kumar, S. (2026). Optimizing air pollution prediction with random forest algorithm. *Aerosol Science and Engineering*, 10, 208–221. <https://doi.org/10.1007/s41810-025-00292-6>
- Solci, C. C., Reisen, V. A., & Rodrigues, P. C. (2023). Robust local bootstrap for weakly stationary time series in the presence of additive outliers. *Stochastic Environmental Research and Risk Assessment*, 37, 2977–2992. <https://doi.org/10.1007/s00477-023-02430-3>

- Taylor, S. J., & Letham, B. (2017). Forecasting at scale. *PeerJ Preprints*, 5, Article e3190v2. <https://doi.org/10.7287/peerj.preprints.3190v2>
- Wen, T., Liu, Y., Bai, Y., & Liu, H. (2023). Modeling and forecasting CO₂ emissions in China and its regions using a novel ARIMALSTM model. *Heliyon*, 9(11), Article e21241. <https://www.sciencedirect.com/science/article/pii/S2405844023084499>
- World Bank. (n. d.). *World Development Indicators* [dataset]. <https://datacatalog.worldbank.org/search/dataset/0037712>
- Wu, S., Feng, Q., Du, Y., & Li, X. (2011). Artificial neural network models for daily PM10 air pollution index prediction in the urban area of Wuhan, China. *Environmental Engineering Science*, 28(5), 357–363. <https://doi.org/10.1089/ees.2010.0219>
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
- Zhang, Y., Wang, J., Chen, R., Zhang, S., Liao, X., & Xie, J. (2024). Complementary ARI-MA and XGBoost carbon emission timing forecasting methods for different terminal energy consumption industries. In *Fifth International Conference on Green Energy, Environment, and Sustainable Development (GEESD 2024)* (Vol. 13279, Article 1327946). SPIE Digital Library. <https://doi.org/10.1117/12.3044690>
- Zheng, Y., Liu, F., & Hsieh, H. P. (2013). U-Air: When urban air quality inference meets big data. In R. Parekh, J. He, D. S. Inderjit, P. Bradley, Y. Koren, R. Ghani, T. E. Senator, R. L. Grossman, & R. Uthrusamy (Eds.), *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1436–1444). Association for Computing Machinery. <https://doi.org/10.1145/2487575.2488188>
- Zhu, S., Lian, X., Wei, L., Che, J., Shen, X., Yang, L., Qiu, X., Liu, X., Gao, W., Ren, X., & Li, J. (2018). PM_{2.5} forecasting using SVR with PSO-GSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors. *Atmospheric Environment*, 183, 20–32. <https://doi.org/10.1016/j.atmosenv.2018.04.004>
- Zhytkevych, O., & Brochado, A. (2022). Modeling national decarbonization capabilities using Kohonen maps. *Neuro-Fuzzy Modeling Techniques in Economics*, 11, 3–24. <http://doi.org/10.33111/nfmte.2022.003>
- Zhytkevych, O., Matviychuk, A., & Kmytiuk, T. (2023). Modeling Nations' Decarbonisation Potential. In F. Ortiz-Rodríguez, S. Tiwari, P. Usoro Usip, & R. Palma (Eds.), *Electronic Governance with Emerging Technologies. EGETC 2023. Communications in Computer and Information Science* (Vol. 1888, pp. 60–77). Springer, Cham. https://doi.org/10.1007/978-3-031-43940-7_6

Додаток А

Таблиця А1

Перелік показників у контексті декарбонізації та сукупність відповідних дослідницьких гіпотез

Категорія	Показник для країни	Джерело	Очікуваний ефект	Дослідницька гіпотеза
Енергетика	Енергоємність ВВП за споживанням (x1)	EnerData, 2024	Позитивний	H0a: $\beta_1 < 0$ H1a: $\beta_1 > 0$
	Внутрішнє споживання нафтопродуктів (x2) Позитивний		Позитивний	H0b: $\beta_2 < 0$ H1b: $\beta_2 > 0$
	Внутрішнє споживання природного газу (x3)		Позитивний	H0c: $\beta_3 < 0$ H1c: $\beta_3 > 0$
	Внутрішнє споживання електроенергії (x4)		Змішаний: залежить від структури енергетичного балансу; зростання частки чистої енергії має сприятливий ефект, тоді як переважання викопного палива — протилежний	H0d: $\beta_4 = 0$ H1d: $\beta_4 \neq 0$
	Частка відновлюваних джерел в електрифікації (x5)		Негативний	H0e: $\beta_5 > 0$ H1e: $\beta_5 < 0$
	Частка вітрової та сонячної енергії в електрифікації (x6)		Негативний	H0f: $\beta_6 > 0$ H1f: $\beta_6 < 0$
	Електрифікація (x7)		Змішаний: залежить від структури енергетичного балансу; зростання частки чистої енергії має сприятливий ефект, тоді як переважання викопного палива — протилежний	H0g: $\beta_7 = 0$ H1g: $\beta_7 \neq 0$
	Внутрішнє споживання вугілля та лігніту (x8)		Позитивний	H0h: $\beta_8 < 0$ H1h: $\beta_8 > 0$
Демографія	Зростання міського населення (річне, %) (x9)	World Bank, 2024	Змішаний: може поліпшуватися завдяки ефективності, але погіршуватися через зростання попиту	H0i: $\beta_9 = 0$ H1i: $\beta_9 \neq 0$
Економіка	Зростання ВВП на душу населення (річне, %) (x10)		Змішаний: вищі доходи можуть фінансувати чисті технології, але водночас підвищувати споживання	H0j: $\beta_{10} = 0$ H1j: $\beta_{10} \neq 0$
Ризик	Індекс геополітичного ризику (GPR) (x11)	Geopolitical Risk Index, 2025	Змішаний: зростання геополітичного ризику може привести до зниження споживання енергії, хоча і знизяться витрати на енергоефективність, що може призвести до більших викидів	H0k: $\beta_{11} = 0$ H1k: $\beta_{11} \neq 0$
Глобалізація	Індекс глобалізації (x12)		Змішаний: може сприяти трансферу чистих технологій, але також збільшувати вуглецевий слід	H0l: $\beta_{12} = 0$ H1l: $\beta_{12} \neq 0$
Довкілля	Середній коефіцієнт викидів CO ₂ (x13)	EnerData, 2024	Цільова змінна	Відсутнє значення

Джерело: розроблено авторами.

Додаток В

Таблиця В1

Описова статистика нормалізованих показників

Показник	Середнє	SD	Медіана
x1	0.3423	0.2340	0.2964
x2	0.0936	0.1799	0.0328
x3	0.0830	0.1572	0.0438
x4	0.0534	0.1316	0.0166
x5	0.2954	0.2517	0.2010
x6	0.1405	0.1650	0.0851
x7	0.5313	0.1861	0.5489
x8	0.6806	0.0952	0.6808
x9	0.6415	0.0961	0.6456
x10	0.0342	0.1270	0.0036
x11	0.4029	0.1472	0.3902
x12	0.4230	0.2534	0.3868
x13	0.5202	0.1836	0.4973

Джерело: розраховано авторами.

Додаток С

Таблиця С1

Результати кластеризації

Кластери	Країни
Кластер 0	Канада (2014–2023), Нова Зеландія (2014–2023), Норвегія (2014–2023), Швеція (2014–2023)
Кластер 1	Австралія (2014–2022), Бельгія (2014–2021), Велика Британія (2014–2016), Малайзія (2016–2023), Нідерланди (2014–2019), Німеччина (2014–2015), Об'єднані Арабські Емірати (2014–2023), Південна Корея (2014–2023), Польща (2014–2022), Сінгапур (2014–2023), Сполучені Штати (2014–2023), Франція (2014–2023), Чехія (2014–2023), Японія (2014–2023)
Кластер 2	Бразилія (2014–2023), Колумбія (2014–2023)
Кластер 3	Іран (2014–2023), Казахстан (2014–2023), Китай (2014–2023), Південна Африка (2014–2023), Росія (2014–2023), Саудівська Аравія (2014–2023), Узбекистан (2014–2022), Україна (2014–2018, 2022–2023)
Кластер 4	Азербайджан (2014–2023), Алжир (2014–2023), В'єтнам (2014–2020), Єгипет (2014–2023), Індія (2014–2023), Індонезія (2014–2023), Малайзія (2014–2015), Мексика (2014–2019, 2021–2023), Нігерія (2014–2023), Таїланд (2014–2023), Туреччина (2014–2023), Узбекистан (2023), Україна (2019–2021)
Кластер 5	Австралія (2023), Бельгія (2020, 2022–2023), Велика Британія (2017–2023), Данія (2014–2023), Іспанія (2014–2018, 2020–2023), Італія (2023), Нідерланди (2020–2023), Німеччина (2019–2023), Португалія (2014–2023), Чилі (2022–2023)
Кластер 6	Аргентина (2014–2023), В'єтнам (2021–2023), Іспанія (2019), Італія (2014–2022), Мексика (2020), Німеччина (2016–2018), Польща (2023), Румунія (2014–2023), Чилі (2014–2021)

Джерело: розроблено авторами.

Додаток D

Фрагмент скрипту:

```
# 3. Define search space for hidden layers (start from 12)
layer1 = 12 # start from 12 neurons
layer2_options = range(3, 10) # second layer: 3-9 neurons
layer3_options = range(2, 7) # third layer: 2-6 neurons
layer4_options = [1] # output layer
results = []

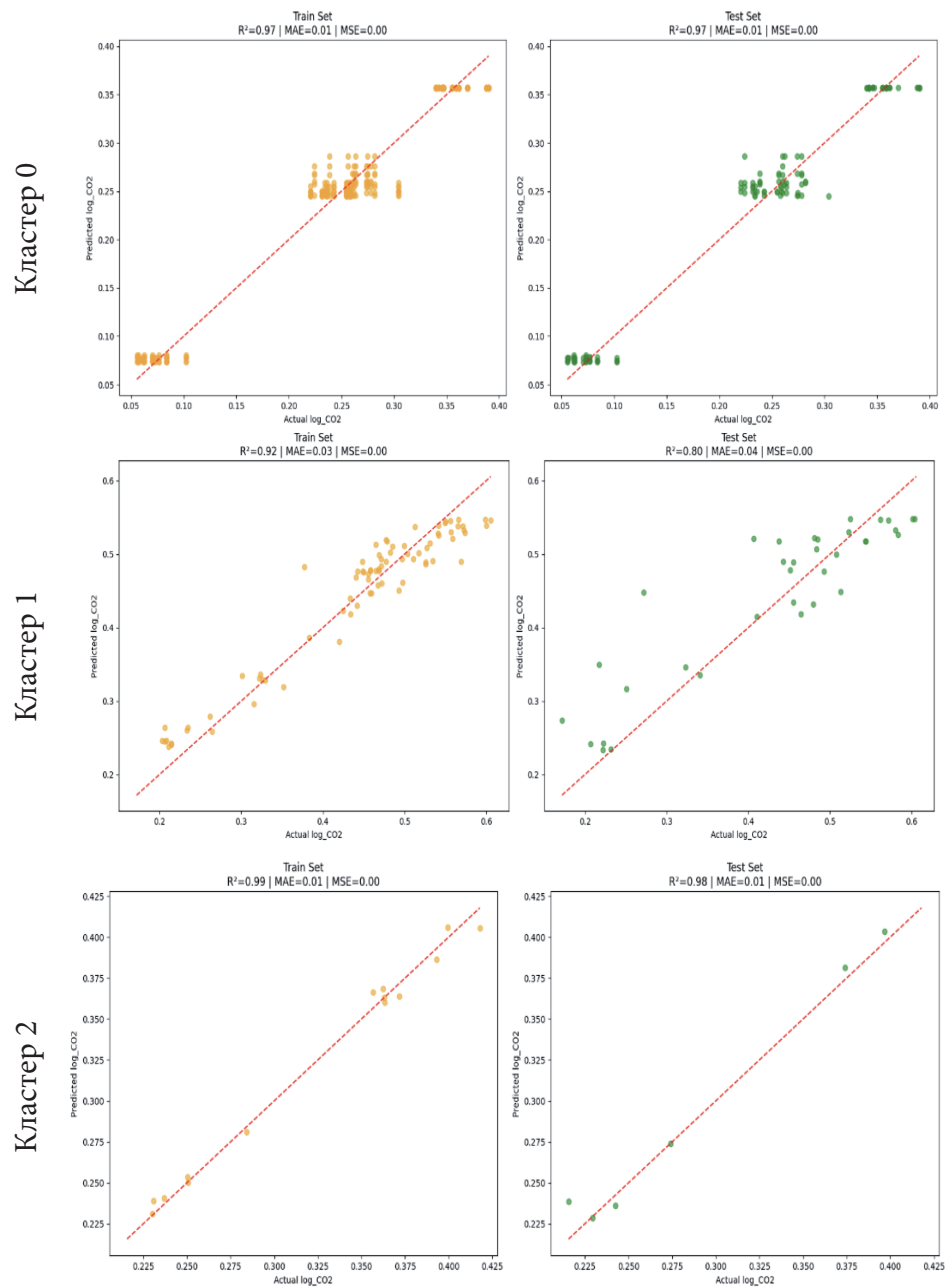
#4. Train and evaluate MLPs for all layer combinations
for l2 in layer2_options:
    for l3 in layer3_options:
        for l4 in layer4_options:
            arch = (layer1, l2, l3, l4)
            mlp = Pipeline([
                ("scale", StandardScaler()),
                ("mlp", MLPRegressor(
                    hidden_layer_sizes=arch,
                    activation="logistic",
                    solver="adam",
                    max_iter=3000,
                    early_stopping=True,
                    n_iter_no_change=30,
                    random_state=42,
                    learning_rate="adaptive",
                    learning_rate_init=0.001
                ))
            ])
            mlp.fit(X_train, y_train)
            y_pred_test = mlp.predict(X_test)
            results.append({
                "Architecture": arch,
                "R2_Test": r2_score(y_test, y_pred_test),
                "MAE_Test": mean_absolute_error(y_test, y_pred_test),
                "MSE_Test": mean_squared_error(y_test, y_pred_test)
            })

# 5. Select the best architecture
results_df = pd.DataFrame(results)
best_idx = results_df["R2_Test"].idxmax()
best_model = results_df.iloc[best_idx]
best_arch = best_model["Architecture"]
print("Best Model")
print(best_model.round(4))
print(f"\n Optimal architecture (calculated): {best_arch}")
```

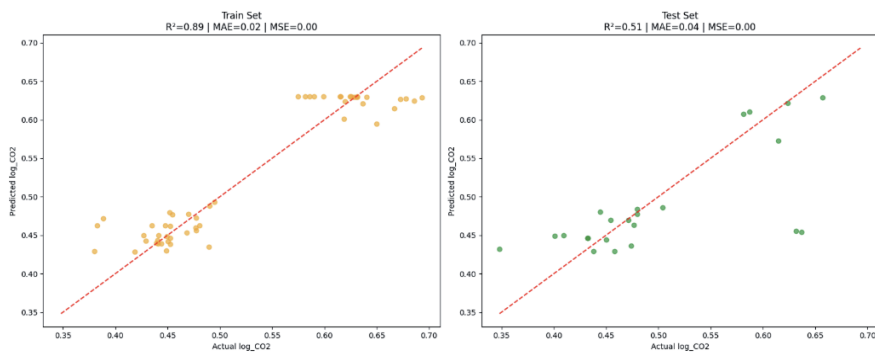
Додаток Е

Рисунок Е1

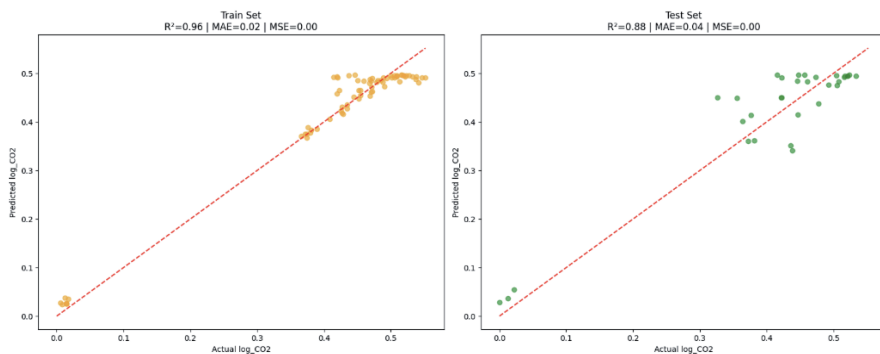
Оцінювання точності прогнозування для 7 кластерів на основі MLP з оптимальною архітектурою для кожного кластера



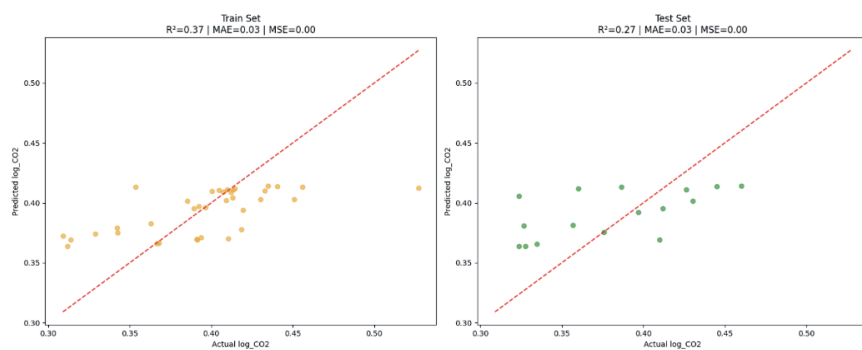
Кластер 3



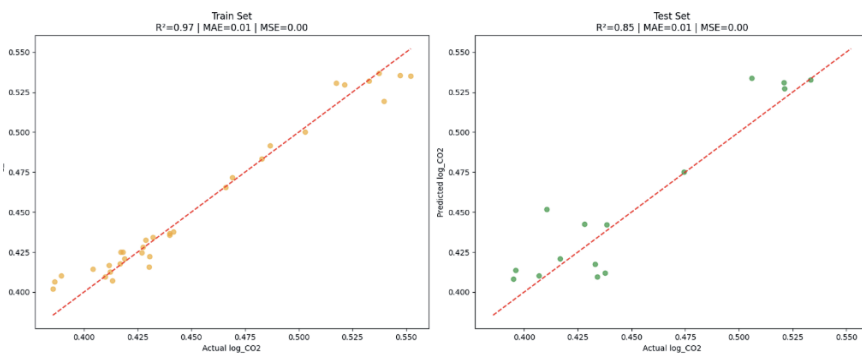
Кластер 4



Кластер 5



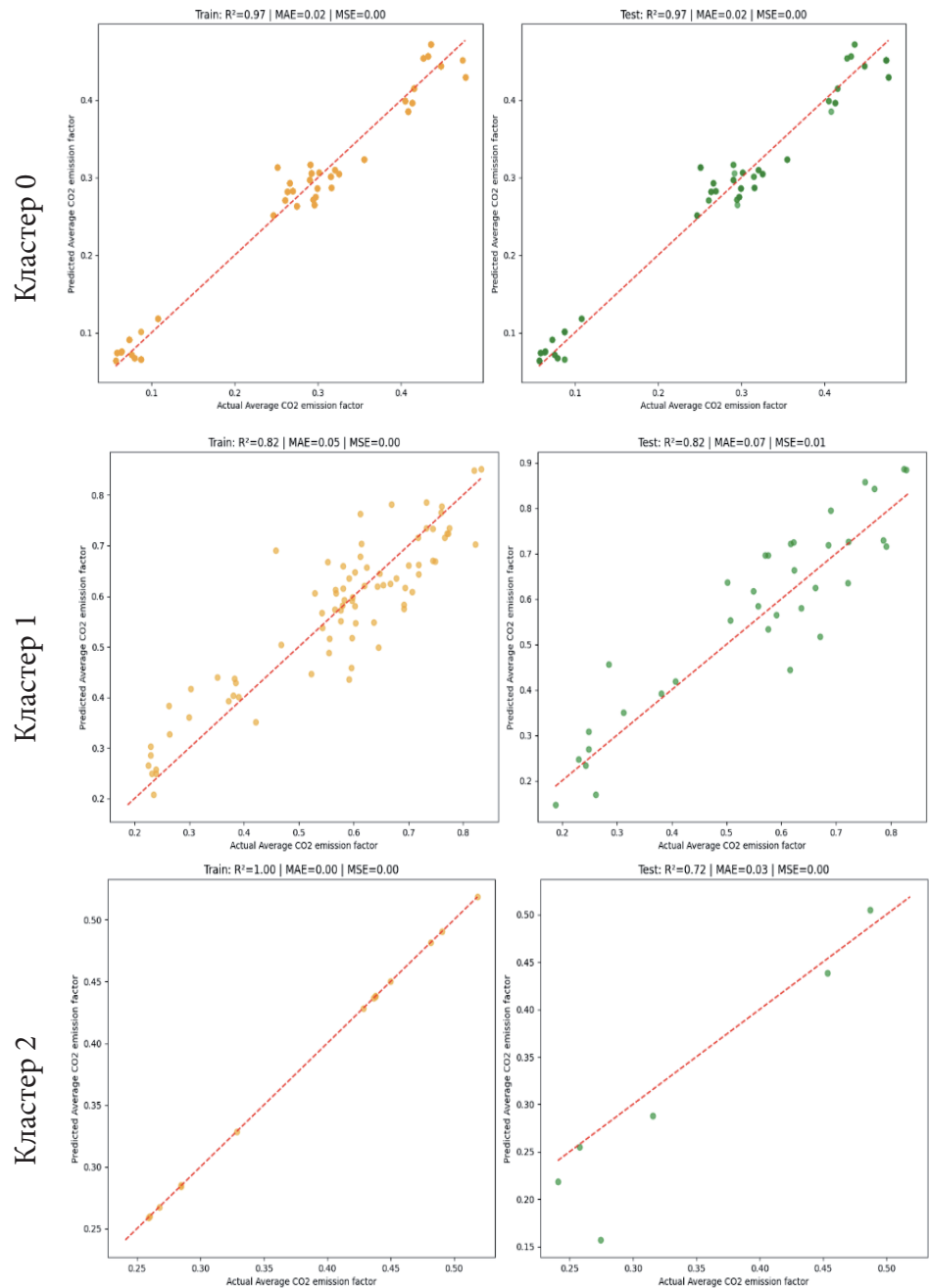
Кластер 6

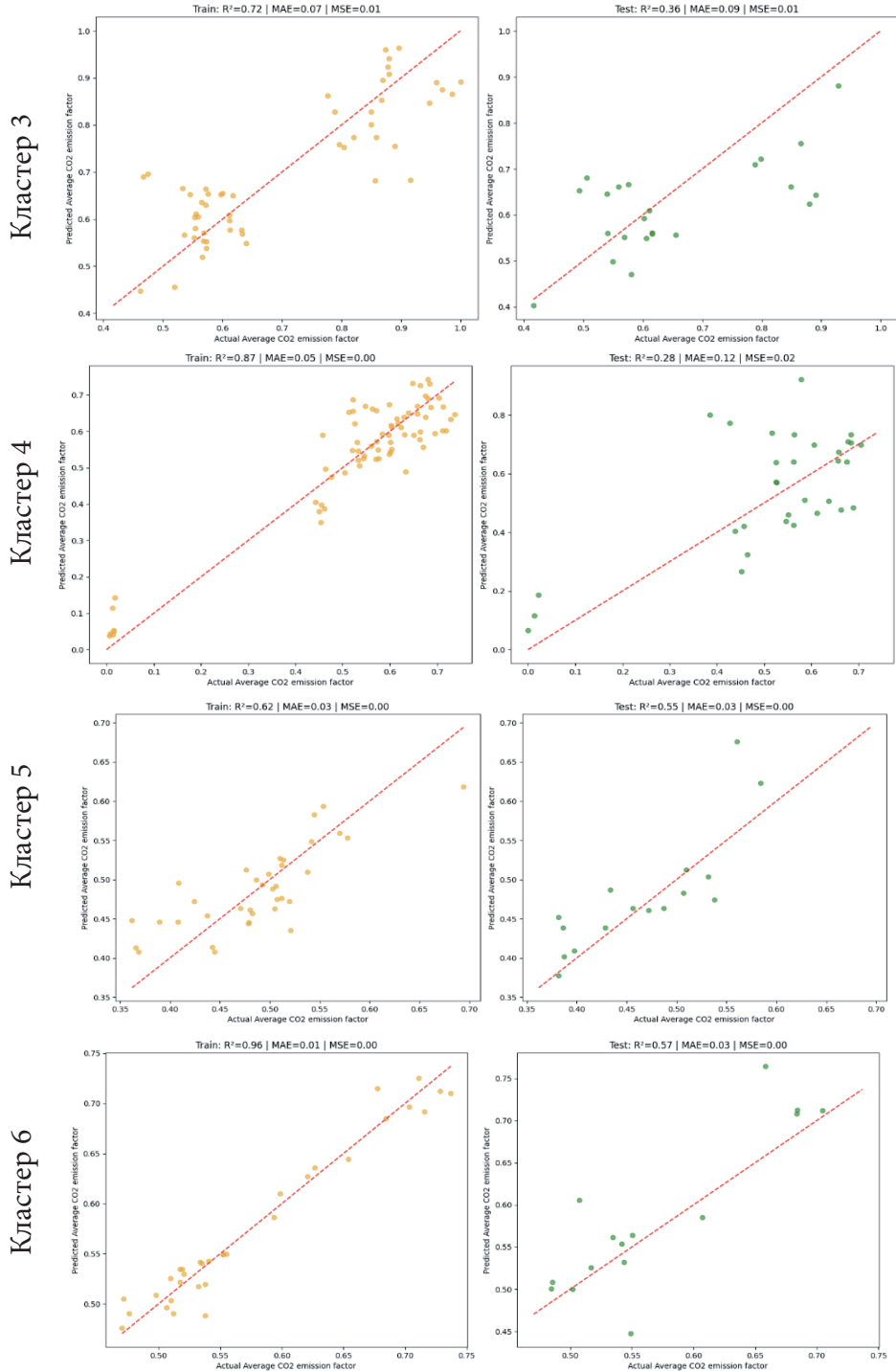


Джерело: розроблено авторами.

Рисунок E2

Оцінювання точності прогнозування для 7 кластерів на основі моделей лінійної регресії





Джерело: розроблено авторами.

Таблиця Е1

Порівняння результативності моделей LR та MLP для 7 кластерів

Модель	Кластери							Середнє
LR	0	1	2	3	4	5	6	
R ² (Навчання)	0,97	0,82	1,00	0,72	0,87	0,62	0,96	0,85
MAE (Навчання)	0,02	0,05	1×10 ⁻⁴	0,07	0,05	0,03	0,01	0,03
MSE (Навчання)	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	0,01	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴
R ² (Тест)	0,97	0,82	0,72	0,36	0,28	0,55	0,57	0,61
MAE (Тест)	0,02	0,07	0,03	0,09	0,12	0,03	0,03	0,06
MSE (Тест)	1×10 ⁻⁴	0,01	1×10 ⁻⁴	0,01	0,02	1×10 ⁻⁴	1×10 ⁻⁴	0,01
MLP	0	1	2	3	4	5	6	
R ² (Навчання)	0,97	0,92	0,99	0,89	0,96	0,37	0,97	0,87
MAE (Навчання)	0,01	0,03	0,01	0,02	0,02	0,03	0,01	0,02
MSE (Навчання)	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴
R ² (Тест)	0,97	0,80	0,98	0,51	0,88	0,27	0,85	0,75
MAE (Тест)	0,01	0,04	0,01	0,04	0,04	0,03	0,01	0,03
MSE (Тест)	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴

Джерело: розроблено авторами.

Додаток F

Таблиця F1

Рівняння лінійної регресії (LR) для 7 кластерів

Кластер	Рівняння лінійної регресії
Кластер 0	$y = 0,173 - 0,230 \cdot \text{Енергоємність ВВП за споживанням} - 0,725 \cdot \text{Внутрішнє споживання нафтопродуктів} + 6,772 \cdot \text{Внутрішнє споживання природного газу} - 15,248 \cdot \text{Внутрішнє споживання електроенергії} + 0,378 \cdot \text{Частка відновлюваних джерел енергії в енергетиці} - 0,382 \cdot \text{Частка вітрової та сонячної енергії в енергетиці} + 0,160 \cdot \text{Приріст міського населення (у \% на рік)} - 0,043 \cdot \text{Зростання ВВП на душу населення (у річному вимірі, \%)} + 43,494 \cdot \text{Внутрішнє споживання кам'яного та бурого вугілля} - 0,111 \cdot \text{Електрифікація} - 0,111 \cdot \text{Індекс геополітичних ризиків} + 0,007 \cdot \text{Індекс глобалізації}$
Кластер 1	$y = -0,033 + 0,011 \cdot \text{Енергоємність ВВП за споживанням} - 1,358 \cdot \text{Внутрішнє споживання нафтопродуктів} + 3,148 \cdot \text{Внутрішнє споживання природного газу} - 5,318 \cdot \text{Внутрішнє споживання електроенергії} - 1,502 \cdot \text{Частка відновлюваних джерел енергії в енергетиці} + 1,331 \cdot \text{Частка вітрової та сонячної енергії в енергетиці} + 0,171 \cdot \text{Приріст міського населення (у \% на рік)} - 0,168 \cdot \text{Зростання ВВП на душу населення (у річному вимірі, \%)} + 7,994 \cdot \text{Внутрішнє споживання кам'яного та бурого вугілля} + 0,319 \cdot \text{Електрифікація} + 0,828 \cdot \text{Індекс геополітичних ризиків} + 1,000 \cdot \text{Індекс глобалізації}$
Кластер 2	$y = 2,170 - 2,622 \cdot \text{Енергоємність ВВП за споживанням} + 5,064 \cdot \text{Внутрішнє споживання нафтопродуктів} - 15,148 \cdot \text{Внутрішнє споживання природного газу} - 4,122 \cdot \text{Внутрішнє споживання електроенергії} - 0,575 \cdot \text{Частка відновлюваних джерел енергії в енергетиці} + 0,622 \cdot \text{Частка вітрової та сонячної енергії в енергетиці} - 0,922 \cdot \text{Приріст міського населення (у \% на рік)} - 0,073 \cdot \text{Зростання ВВП на душу населення (у річному вимірі, \%)} + 50,395 \cdot \text{Внутрішнє споживання кам'яного та бурого вугілля} - 0,855 \cdot \text{Електрифікація} - 0,374 \cdot \text{Індекс геополітичних ризиків} + 0,377 \cdot \text{Індекс глобалізації}$
Кластер 3	$y = -0,013 + 0,254 \cdot \text{Енергоємність ВВП за споживанням} - 0,161 \cdot \text{Внутрішнє споживання нафтопродуктів} + 0,090 \cdot \text{Внутрішнє споживання природного газу} - 1,340 \cdot \text{Внутрішнє споживання електроенергії} + 0,127 \cdot \text{Частка відновлюваних джерел енергії в енергетиці} + 1,290 \cdot \text{Частка вітрової та сонячної енергії в енергетиці} + 0,418 \cdot \text{Приріст міського населення (у \% на рік)} + 0,134 \cdot \text{Зростання ВВП на душу населення (у річному вимірі, \%)} + 1,147 \cdot \text{Внутрішнє споживання кам'яного та бурого вугілля} + 0,660 \cdot \text{Електрифікація} - 0,111 \cdot \text{Індекс геополітичних ризиків} - 0,065 \cdot \text{Індекс глобалізації}$
Кластер 4	$y = -0,329 + 0,132 \cdot \text{Енергоємність ВВП за споживанням} + 1,750 \cdot \text{Внутрішнє споживання нафтопродуктів} - 0,074 \cdot \text{Внутрішнє споживання природного газу} - 20,538 \cdot \text{Внутрішнє споживання електроенергії} - 0,903 \cdot \text{Частка відновлюваних джерел енергії в енергетиці} + 2,914 \cdot \text{Частка вітрової та сонячної енергії в енергетиці} + 0,741 \cdot \text{Приріст міського населення (у \% на рік)} - 0,254 \cdot \text{Зростання ВВП на душу населення (у річному вимірі, \%)} + 9,946 \cdot \text{Внутрішнє споживання кам'яного та бурого вугілля} + 1,728 \cdot \text{Електрифікація} - 0,612 \cdot \text{Індекс геополітичних ризиків} + 0,751 \cdot \text{Індекс глобалізації}$
Кластер 5	$y = 0,200 - 0,178 \cdot \text{Енергоємність ВВП за споживанням} + 4,200 \cdot \text{Внутрішнє споживання нафтопродуктів} + 3,386 \cdot \text{Внутрішнє споживання природного газу} - 17,846 \cdot \text{Внутрішнє споживання електроенергії} - 0,546 \cdot \text{Частка відновлюваних джерел енергії в енергетиці} + 0,303 \cdot \text{Частка вітрової та сонячної енергії в енергетиці} + 0,054 \cdot \text{Приріст міського населення (у \% на рік)} - 0,051 \cdot \text{Зростання ВВП на душу населення (у річному вимірі, \%)} + 10,619 \cdot \text{Внутрішнє споживання кам'яного та бурого вугілля} + 0,664 \cdot \text{Електрифікація} + 0,132 \cdot \text{Індекс геополітичних ризиків} + 0,329 \cdot \text{Індекс глобалізації}$
Кластер 6	$y = 0,687 + 0,428 \cdot \text{Енергоємність ВВП за споживанням} + 1,028 \cdot \text{Внутрішнє споживання нафтопродуктів} + 0,955 \cdot \text{Внутрішнє споживання природного газу} - 6,813 \cdot \text{Внутрішнє споживання електроенергії} - 0,234 \cdot \text{Частка відновлюваних джерел енергії в енергетиці} - 0,199 \cdot \text{Частка вітрової та сонячної енергії в енергетиці} - 0,070 \cdot \text{Приріст міського населення (у \% на рік)} - 0,004 \cdot \text{Зростання ВВП на душу населення (у річному вимірі, \%)} + 4,097 \cdot \text{Внутрішнє споживання кам'яного та бурого вугілля} + 0,363 \cdot \text{Електрифікація} + 0,271 \cdot \text{Індекс геополітичних ризиків} - 0,473 \cdot \text{Індекс глобалізації}$

Джерело: розроблено авторами.

Додаток G

Таблиця G1

Узагальнення значущих показників за кластерами та рекомендованих методів моделювання

Кластер	Значущі показники	Рекомендований метод моделювання
0	Енергоємність ВВП, Споживання електроенергії	Лінійна регресія (LR)
1	Споживання електроенергії, Споживання кам'яного та бурого вугілля, Природний газ	Лінійна регресія (LR)
2	Частка відновлюваних джерел енергії, Вітрова та сонячна генерація	Нелінійні моделі (MLP)
3	Зростання чисельності міського населення, Зростання ВВП на душу населення	Нелінійні моделі (MLP)
4	Електрифікація, Індекс глобалізації	Нелінійні моделі (MLP)
5	Геополітичний ризик, Інтенсивність ВВП	Лінійна регресія (LR)
6	Споживання кам'яного та бурого вугілля, Електрифікація	Нелінійні моделі (MLP)

Джерело: розроблено авторами.

Додаток Н

Таблиця Н1

Перевірка гіпотез на наборах даних для кожного із 7 кластерів

Показник		x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12
Очікуваний знак (H1)		>0	>0	>0	Будь-який	<0	<0	Будь-який	>0	Будь-який	Будь-який	Будь-який	Будь-який
Кластер 0	β	-0.22	-0.61	6.74	-15.42	0.37	-0.37	0.16	-0.02	44.80	-0.09	-0.11	0.04
	p	1×10 ⁻⁴	0.41	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	0.30	1×10 ⁻⁴	0.12	1×10 ⁻⁴	0.38
Кластер 1	β	0.08	-1.92	3.05	-4.10	-1.37	1.31	0.14	-0.07	8.48	0.37	0.74	0.82
	p	0.34	0.01	1×10 ⁻⁴	0.01	1×10 ⁻⁴	1×10 ⁻⁴	0.17	0.53	1×10 ⁻⁴	0.01	1×10 ⁻⁴	1×10 ⁻⁴
Кластер 2	β	-1.47	1.56	-1.05	-4.53	-0.20	0.33	-0.58	-0.07	34.82	-0.95	-0.10	0.12
	p	0.01	0.01	0.62	0.08	0.10	0.03	0.06	0.06	0.27	1×10 ⁻⁴	0.51	0.20
Кластер 3	β	0.26	0.22	-0.07	-1.65	0.47	1.65	0.43	0.16	1.06	0.45	-0.15	-0.04
	p	0.17	0.74	0.81	0.01	0.37	0.01	0.00	0.20	0.08	0.02	0.37	0.80
Кластер 4	β	-0.06	0.36	1.94	-14.12	-0.55	1.43	0.54	-0.05	8.28	1.34	-0.37	0.44
	p	0.59	0.76	0.12	0.01	0.01	1×10 ⁻⁴	0.05	0.73	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴	1×10 ⁻⁴
Кластер 5	β	-0.05	1.39	3.18	-11.09	-0.58	0.34	-0.08	0.03	8.31	0.72	0.05	0.33
	p	0.82	0.69	0.03	0.23	1×10 ⁻⁴	0.03	0.79	0.61	1×10 ⁻⁴	0.05	0.71	0.10
Кластер 6	β	0.37	0.51	-0.82	0.19	-0.05	-0.11	0.16	0.03	1.23	-0.19	0.25	-0.27
	p	0.08	0.58	0.54	0.96	0.72	0.23	0.30	0.49	0.59	0.47	1×10 ⁻⁴	0.01
Висновки		Значущий ефект, гіпотеза H1 підтверджено для кластера 6	Значущий ефект, гіпотеза H1 підтверджено для кластера 2	Значущий ефект, гіпотеза H1 підтверджено для кластерів 0, 1, 5	Значущий ефект, гіпотеза H1 підтверджено для кластерів 0-4	Значущий ефект, гіпотеза H1 підтверджено для кластерів 1, 2, 4, 5	Значущий ефект, H1 підтверджено для кластера 0	Значущий ефект, H1 підтверджено для кластерів 0, 2, 3, 4	Значущий ефект, H1 підтверджено для кластерів 0, 2	Значущий ефект, H1 підтверджено для кластерів 0, 1, 3, 4, 5	Значущий ефект, H1 підтверджено для кластерів 0, 1, 2, 4, 5	Значущий ефект, H1 підтверджено для кластерів 0, 1, 4, 6	Значущий ефект, H1 підтверджено для кластерів 0, 1, 4, 5, 6

Джерело: розроблено авторами.

Отримано: 5 квітня 2026 р.
Рецензовано: 6 серпня 2026 р.
Рекомендовано до друку: 16 вересня 2026 р.