

8. Stebila, D., Fluhrer, S., & Gueron, S. (2025). Hybrid key exchange in Transport Layer Security (TLS) 1.3 (RFC 9650) [Електронний ресурс]. Internet Engineering Task Force. Режим доступу: <https://www.rfc-editor.org/rfc/rfc9650>
9. M. B., Schinzel, J., Lange, T., & Bernstein, D. J. (2025). Hybrid post-quantum key exchange in Secure Shell (SSH) (RFC 9710) [Електронний ресурс]. Internet Engineering Task Force. Режим доступу: <https://www.rfc-editor.org/rfc/rfc9710>.
10. Cybersecurity and Infrastructure Security Agency. (2024). Link layer authentication and link layer encryption: Are you really secure? [Електронний ресурс]. Режим доступу: https://www.cisa.gov/sites/default/files/2024-09/24_0828_s-n_LLA_LLE_are_you_really_secure_2022_final_508C.pdf.

Надійшла 14.10.2025

УДК 004.85:330.322

DOI: 10.31673/2409-7292.2025.041210

Криворучко В. Ф., Садовенко В. С.

ІНТЕГРОВАНІЙ ПІДХІД ДО АНАЛІЗУ ІНВЕСТИЦІЙ У ЦИФРОВУ ІНФРАСТРУКТУРУ НА ОСНОВІ КЛАСТЕРИЗАЦІЇ ТА РЕГРЕСІЙНИХ МОДЕЛЕЙ

Стаття присвячена розробці інтегрованого підходу до аналізу інвестицій у цифрову інфраструктуру на основі поєднання алгоритмів кластеризації та методів регресійного моделювання. Актуальність дослідження зумовлена зростаючою роллю цифрової інфраструктури у забезпеченні економічної конкурентоспроможності регіонів та необхідністю оптимізації інвестиційних рішень в умовах обмежених ресурсів. У роботі використано алгоритми K-Means і DBSCAN для сегментації регіонів за показниками цифрової зрілості, рівнем ІТ-зайнятості, інтернет-покриття та доступності цифрових сервісів. Для кожного отриманого кластера побудовано окрему множину лінійну регресійну модель, що дає змогу оцінити вплив ключових цифрових факторів на обсяг інвестицій. Якість моделей оцінено за метриками R^2 , RMSE та MAE із застосуванням k-fold cross-validation. Інтеграція результатів кластеризації та регресійного аналізу дозволила сформувати матрицю аналітичної оцінки, яка порівнює фактичні та прогнозовані інвестиції і визначає пріоритетні напрями інвестиційної політики. Запропонований підхід забезпечує підвищення обґрунтованості управлінських рішень, дозволяє виявляти ефективні та неефективні регіони, а також формувати рекомендації щодо оптимізації фінансування цифрової інфраструктури. Отримані результати можуть бути використані органами влади, аналітичними центрами та ІКТ-компаніями для підтримки стратегічного планування цифрового розвитку.

Ключові слова: цифрова інфраструктура; інвестиції; кластеризація; K-Means; DBSCAN; регресійний аналіз; цифрова економіка; машинне навчання.

Вступ

Цифрова трансформація стала одним із ключових драйверів економічного розвитку у ХХІ столітті. Розвиток цифрової інфраструктури (телекомунікаційних мереж, дата-центрів, хмарних сервісів, систем кібербезпеки) визначає конкурентоспроможність держави на глобальному ринку. Водночас ефективність інвестицій у цю сферу залишається неоднорідною, що зумовлює необхідність використання нових аналітичних інструментів для підтримки управлінських рішень.

Традиційні економетричні методи аналізу, зокрема класичні моделі регресії або кореляційного аналізу, не завжди здатні адекватно відобразити складні, нелінійні взаємозв'язки між показниками цифрового розвитку та обсягами інвестицій. Натомість алгоритми машинного навчання, особливо методи кластеризації та регресійного аналізу, відкривають можливість для глибшого виявлення закономірностей у багатовимірних економічних даних. Серед сучасних підходів до кластерного аналізу особливе місце посідають алгоритми K-Means і DBSCAN, які дозволяють групувати об'єкти за подібністю характеристик або виділяти аномальні спостереження у даних. Поєднання цих алгоритмів із методами множинної лінійної регресії створює основу для побудови комплексних моделей прогнозування та оптимізації інвестиційних рішень у сфері цифрової інфраструктури.

Актуальність дослідження зумовлена необхідністю підвищення ефективності державних і приватних інвестицій у цифровий сектор, особливо в умовах обмежених ресурсів та швидкої зміни технологічних пріоритетів. Водночас практична реалізація аналітичних підходів, що

базуються на машинному навчанні, залишається недостатньо розробленою в контексті національних стратегій цифрової економіки України.

Таким чином, доцільним є створення аналітичної моделі, яка поєднає кластеризацію та регресійний аналіз для визначення груп регіонів або секторів із подібними характеристиками цифрового розвитку та прогнозування оптимального рівня інвестицій.

Огляд літератури

Питання управління інвестиціями в цифрову інфраструктуру напряму пов'язане з трансформацією економіки під впливом цифрових технологій. Brynjolfsson та McAfee розглядають цифрові технології як фундаментальний фактор зміни продуктивності, структури ринку, підкреслюючи роль інфраструктури даних, обчислень та мереж як бази «другої машинної доби» [1]. Це формує загальноекономічний контекст для оцінки інвестицій у цифрову інфраструктуру.

Goldfarb і Tucker у статті «Digital Economics» систематизують вплив зниження вартості зберігання, обробки та передавання даних на ринки, конкуренцію та політику регулювання, окреслюючи методологічні підходи до аналізу цифрових ринків і платформ [2]. Це забезпечує теоретичне підґрунтя для використання кількісних моделей при аналізі інвестиційних рішень у цифровому середовищі. З методологічної точки зору, фундаментом для побудови моделей зв'язку між цифровими індикаторами та інвестиційними показниками залишаються класичні підходи регресійного аналізу, детально представлені у праці Montgomery, Peck та Vining [3]. Вони формують базу для коректної специфікації, оцінювання та валідації множинних регресійних моделей, що використовується в даному дослідженні.

У частині методів кластеризації ключовими є роботи Jain [4] та Ester et al. [5]. Jain узагальнює півстоліття розвитку методів кластеризації та підкреслює роль алгоритму K-Means як базового інструменту для поділу об'єктів за подібністю. Ester та співавтори пропонують алгоритм DBSCAN, орієнтований на щільнісне групування та виявлення аутлайєрів у великих просторових вибірках [5]. Саме поєднання K-Means (для структуризації даних) та DBSCAN (для ідентифікації регіонів/об'єктів з нетиповою інвестиційною динамікою) є методологічно виправданим у контексті даної статті. Сучасні прикладні дослідження демонструють доцільність використання методів машинного навчання для прогнозування економічних показників та підтримки рішень. У роботі Haluzov et al. [6] показано застосування методів data science для прогнозування попиту в ритейлі, що підтверджує ефективність поєднання регресійних і класифікаційних моделей у задачах з високою варіативністю даних. У статті «Порівняльний аналіз методів машинного навчання для побудови прогнозів» [7] показано порівняльний аналіз методів машинного навчання для побудови прогнозів, акцентуючи на залежності точності від вибору алгоритму та метрик якості. У публікації [8] розглянуто інструменти очищення даних для прогнозування інвестицій за допомогою машинного навчання, що безпосередньо корелює з етапом підготовки даних у запропонованій методології. Таким чином, наявні праці формують методологічний каркас для використання кластеризації (K-Means, DBSCAN) і регресійного аналізу в дослідженні цифрової інфраструктури, однак інтегрована модель сегментації регіонів та побудови кластер-специфічних регресій для оптимізації інвестицій залишається недостатньо розробленою. Саме заповнення цієї прогалини і становить наукову новизну даного дослідження.

Мета дослідження

Розробити підхід до аналітичного обґрунтування управлінських рішень щодо інвестицій у цифрову інфраструктуру на основі алгоритмів K-Means, DBSCAN та методів регресійного аналізу.

Метод дослідження

Розроблений аналітичний підхід до управління інвестиціями в цифрову інфраструктуру передбачає інтеграцію методів кластеризації (K-Means, DBSCAN) і регресійного аналізу у єдину модель. Структуру методу представлено на рисунку 1.

У дослідженні використано статистичні дані щодо регіонів України за період 2019—2024 рр., що включають такі показники:

- X_1 — індекс цифрової економіки (CEDI);
- X_2 — частка ІТ-фахівців у структурі зайнятості (%);
- X_3 — рівень проникнення інтернету (% домогосподарств);
- X_4 — кількість цифрових адміністративних сервісів;
- X_5 — обсяг капітальних інвестицій у сектор ІКТ (млн грн);
- Y — цільова змінна — загальний обсяг інвестицій у цифрову інфраструктуру регіону.

Перед проведенням аналітичних процедур дані було очищено від пропусків і нормалізовано методом Min–Max Scaling за формулою:

$$x_{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}}, \quad (1)$$

що забезпечує приведення значень до інтервалу $[0;1]$ та усуває вплив різних масштабів ознак.

Для виявлення структурних подібностей між регіонами застосовано два підходи.



Рис. 1. Блок-схема аналітичного підходу до управління інвестиціями (K-Means + DBSCAN + Regression)

Алгоритм K-Means, який мінімізує функцію сум квадратів відстаней між точками та центроїдами кластерів:

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2, \quad (2)$$

де k — кількість кластерів, μ_i — центр i -го кластера, C_i — множина об'єктів кластера.

Оптимальну кількість кластерів визначено методом «лікоть» (Elbow Method), що базується на аналізі зміни функції $J(k)$.

Алгоритм DBSCAN (Density-Based Spatial Clustering of Applications with Noise), який використовує параметри ε (радіус околу) та minPts (мінімальна кількість сусідів).

Кластер формується навколо точок, для яких кількість об'єктів у радіусі ε не менша за minPts.

Цей метод дозволяє виявляти регіони-аутлайєри з нестандартними характеристиками інвестицій або цифрової активності.

Для візуалізації результатів кластеризації використано метод аналізу головних компонент (PCA), що знижує розмірність простору до двох координат і дозволяє графічно відобразити межі кластерів.

Для кожного кластера, отриманого на попередньому етапі, побудовано множинну лінійну регресійну модель, яка описує вплив цифрових факторів на інвестиційну активність:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i, \quad (3)$$

де Y_i — прогнозований рівень інвестицій у регіоні i ; β_j — коефіцієнти регресії, що оцінюють силу впливу кожного чинника; ε_i — стохастичний залишок.

Параметри оцінено методом найменших квадратів (OLS), а достовірність моделі перевірено за допомогою:

- коефіцієнта детермінації R^2 ;
- середньоквадратичної похибки (RMSE);
- середньої абсолютної похибки (MAE).

Для запобігання перенаванчання застосовано метод k-fold cross-validation із параметром $k = 5$, що передбачає послідовне навчання моделі на 80 % даних і тестування на решті 20 %. Такий підхід забезпечує стійкість та узагальнювальну здатність моделі при різних розподілах вибірки.

Результати кластеризації та регресійного аналізу інтегровано у матрицю аналітичної оцінки, яка дозволяє:

- виявити регіони з найбільш ефективним використанням цифрових ресурсів;
- визначити потенційні напрямки оптимізації інвестицій;
- підтримати формування портфеля інвестиційних рішень у межах державних і корпоративних стратегій цифровізації.

Інтерпретація здійснювалася на основі порівняльних метрик моделей, а також аналізу середніх значень цифрових індикаторів у межах кожного кластера.

Результати дослідження

За результатами експериментального моделювання реалізовано інтегрований підхід, що дозволив провести кластеризацію та подальший регресійний аналіз. Попередньо дані пройшли нормалізацію та перевірку на мультиколінеарність, що забезпечило коректність подальшого моделювання.

Аналітичний підхід до управління інвестиціями в цифрову інфраструктуру базується на поєднанні кластерного аналізу та множинного регресійного моделювання. Нехай $X = \{x_{ij}\}$, $i=1, \dots, n$; $j=1, \dots, m$ — матриця ознак, що відображає показники цифрового розвитку регіонів, а $Y = \{y_i\}$ — вектор цільових значень (обсяг інвестицій).

Перед моделюванням дані нормалізуються методом Min–Max Scaling, що приводить усі змінні до інтервалу $[0; 1]$, використовуючи формулу (1).

Кластеризація методом K-Means виконується шляхом мінімізації суми квадратів відстаней між об'єктами та центроїдами кластерів за формулою (2).

Для виявлення щільніших структур і аутлайерів додатково використовується алгоритм DBSCAN. Для заданого значення радіуса ε визначається ε -окіл точки x :

$$N_\varepsilon(x) = \{y \in X : \|x - y\| \leq \varepsilon\}, \quad (4)$$

де параметр ε задає радіус околу, а точка x вважається ядром кластера, якщо $|N_\varepsilon(x)| \geq \text{minPts}$.

Після кластеризації для кожного кластера будується модель множинної лінійної регресії, яка описує зв'язок між цифровими показниками та обсягом інвестицій:

$$\hat{y}_i = \beta_0 + \sum_{j=1}^m \beta_j x_{ij} + \varepsilon_i, \quad (5)$$

де β_j — параметри моделі, оцінені методом найменших квадратів (OLS).

Якість побудованих моделей оцінюється за допомогою коефіцієнта детермінації R^2 , середньоквадратичної похибки (RMSE) та середньої абсолютної похибки (MAE):

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_i (y_i - \hat{y}_i)^2}, \quad (7)$$

$$MAE = \frac{1}{n} \sum_i |y_i - \hat{y}_i|. \quad (8)$$

За допомогою алгоритму K-Means на основі цифрових та інвестиційних показників регіони України було поділено на три основні кластери. Вибір $k=3$ підтверджено методом «лікоть» (мінімізація внутрішньокластерної дисперсії). Значення таблиць сформовано за результатами експериментального моделювання на основі узагальненої (синтетичної) вибірки, що не відображає офіційних статистичних даних.

Таблиця 1

Характеристика кластерів регіонів України за рівнем цифрової зрілості та інвестиційної активності

Кластер	Характеристика регіонів	Середній індекс цифровізації	Середній обсяг інвестицій (млн грн)	Типовий приклад регіону
1	Високий рівень цифрового розвитку, розвинена IT-інфраструктура, наявність R&D центрів	0,86	5120	Київ, Львів
2	Середній рівень цифрової готовності, зростаюча частка IT-ринку, обмежене 5G-покриття	0,63	3170	Харків, Одеса
3	Низький рівень цифровізації, слабка інфраструктура та освітній потенціал	0,41	1750	Чернігів, Волинь

Візуалізація результатів у 2D-просторі головних компонент (PCA) показала чітке розділення груп, що підтверджує релевантність кластерної структури. Алгоритм DBSCAN із параметрами $\varepsilon=0,35$, $\text{minPts}=4$ дозволив виявити регіони-аутлайери, які характеризуються високими інвестиціями при низькому рівні цифрової зрілості, що свідчить про неефективність розподілу ресурсів.

Оцінку параметрів здійснено методом найменших квадратів (OLS), а ефективність моделей перевірено за показниками R^2 , RMSE та MAE. Результати подано в таблиці 2.

Таблиця 2

Показники якості регресійних моделей для кожного кластера

Кластер	R^2	RMSE	MAE	Найвпливовіший чинник	Інтерпретація
1	0,89	210,5	165,2	Рівень IT-зайнятості (X_2)	У високорозвинених регіонах основним фактором зростання інвестицій є людський капітал
2	0,81	285,7	240,1	Покриття 4G/5G (X_3)	Регіони середнього рівня розвитку реагують на розширення телеком-інфраструктури
3	0,74	310,9	268,4	Кількість цифрових сервісів (X_4)	У регіонах з низькою цифровізацією інвестиції стимулює державна підтримка сервісів е-урядування

Інтегруючи результати кластеризації та регресійного аналізу, побудовано матрицю аналітичної оцінки, що дозволяє порівняти фактичні та прогнозовані показники інвестицій для кожної групи регіонів і визначити пріоритети управлінських рішень (табл. 3).

Отримані результати підтверджують доцільність поєднання методів кластеризації та регресійного аналізу для формування гнучкої аналітичної системи підтримки управлінських рішень. Кластери високої цифрової зрілості характеризуються стабільними моделями з високим R^2 , що свідчить про передбачуваність інвестиційної поведінки. Кластери середнього рівня демонструють потенціал до зростання за рахунок модернізації мереж і підвищення рівня цифрових послуг. Кластери низької цифровізації вимагають державного стимулювання та залучення приватного капіталу через публічно-приватні партнерства.

Таблиця 3

Матриця аналітичної оцінки ефективності інвестицій
у цифрову інфраструктуру за кластерами

Кластер	Середній фактичний рівень інвестицій (млн грн)	Прогноз за моделлю (Y_{pred})	Відхилення, %	Ефективність використання ресурсів	Рекомендації
1	5120	5050	-1,4	Висока	Підтримувати інноваційні кластери, розвивати R&D
2	3170	3340	+5,4	Середня	Збільшити інвестиції у телекомунікаційну інфраструктуру
3	1750	1620	-7,4	Низька	Спрямувати ресурси на цифрову освіту та розвиток e-сервісів

Таким чином, інтегрований підхід забезпечує не лише кількісну оцінку інвестиційних тенденцій, а й стратегічну аналітику для визначення пріоритетних напрямів фінансування цифрової інфраструктури.

Обговорення результатів.

Отримані результати демонструють, що поєднання методів кластеризації та регресійного аналізу забезпечує комплексний інструментарій для оцінювання інвестицій у цифрову інфраструктуру за умов значної неоднорідності регіонального розвитку. У процесі кластеризації було виявлено суттєві відмінності між групами регіонів, що підтверджує наявність просторової та структурної поляризації цифрової економіки в Україні. За результатами регресійного аналізу встановлено, що домінантні фактори інвестиційної активності відрізняються залежно від типу кластера: у регіонах із високим рівнем цифрової зрілості провідну роль відіграє людський капітал, тоді як у регіонах середнього рівня цифровізації ключовим чинником виступає розвиток телекомунікаційної інфраструктури. Для регіонів із низьким рівнем цифрового розвитку визначальним фактором залишається державна підтримка електронних сервісів та цифрових платформ.

Таке розмежування факторів узгоджується з міжнародними дослідженнями цифрової економіки, де підкреслюється важливість людського капіталу та інноваційної екосистеми для розвинених центрів, а також критичність базової телекомунікаційної інфраструктури для країн, що перебувають на етапі формування цифрової екосистеми. У країнах ЄС та Південно-Східної Азії, наприклад, прослідковується аналогічний перехід: від інфраструктурних інвестицій на ранніх етапах до інвестицій у навички та цифрові сервіси на більш зрілих стадіях. Це підтверджує універсальність встановлених закономірностей та адаптивність запропонованого підходу до ширшого контексту міжнародних цифрових стратегій.

Окремої уваги заслуговує виявлення регіонів-аутлайерів за допомогою алгоритму DBSCAN. Ці регіони характеризуються дисбалансом між рівнем цифрової зрілості та фактичними обсягами інвестицій, що може свідчити про неефективне використання ресурсів, затримку в реалізації інфраструктурних проєктів або відсутність інтегрованого підходу до управління інвестиціями. Для МЦТ такі регіони можуть бути індикаторами необхідності проведення аудиту інвестиційних програм, переорієнтації фінансування або формування цільових цифрових інтервенцій.

Комплексна інтеграція результатів кластеризації та регресійного аналізу у вигляді матриці аналітичної оцінки дозволяє сформулювати практичні рекомендації щодо підвищення ефективності інвестиційної політики. Зокрема, регіони з високим рівнем цифрової зрілості потребують підтримки інноваційних кластерів та R&D-ініціатив, тоді як регіони середнього рівня – державної й приватної координації для розвитку телекомунікаційної інфраструктури. Регіони з низьким рівнем цифровізації мають отримувати підтримку цифрових освітніх

програм та розвиток е-сервісів, що створює передумови для подальшого інвестиційного зростання.

Загалом результати підтверджують, що запропонований підхід сприяє підвищенню точності інвестиційного планування та дозволяє формувати диференційовані рішення з урахуванням реального стану цифрової екосистеми регіонів.

Висновки

У статті запропоновано інтегрований аналітичний підхід до управління інвестиціями в цифрову інфраструктуру, який поєднує алгоритми кластеризації (K-Means, DBSCAN) і методи множинного регресійного аналізу. Такий підхід дозволяє системно оцінювати рівень цифрової зрілості регіонів і визначати оптимальні напрями інвестиційної підтримки.

Проведене дослідження показало, що кластеризація ефективно розподіляє регіони за рівнем цифрового розвитку, виявляючи групи з подібними характеристиками та аутлайєри, що демонструють неефективність використання ресурсів. Подальший регресійний аналіз дозволив кількісно оцінити вплив окремих цифрових показників – рівня ІТ-зайнятості, покриття 4G/5G та кількості е-сервісів – на обсяг інвестицій у кожному кластері.

Отримані результати підтверджують, що у регіонах із високим рівнем цифрової зрілості ключовим драйвером інвестицій виступає людський капітал, а у регіонах середнього розвитку – розбудова телекомунікаційної інфраструктури, у регіонах низької цифровізації – державні цифрові програми та розвиток освітніх ініціатив.

Наукова новизна роботи полягає у побудові комплексної моделі, яка поєднує методи машинного навчання (K-Means, DBSCAN) з економетричними підходами для аналітичного обґрунтування управлінських рішень у сфері цифрових інвестицій.

Практичне значення результатів полягає у можливості використання розробленої матриці аналітичної оцінки для формування регіональних стратегій цифрового розвитку, оптимізації державного фінансування та прогнозування ефективності інвестиційних проєктів.

У подальших дослідженнях доцільно розширити підхід шляхом застосування гібридних ансамблевих моделей (наприклад, Gradient Boosting, Random Forest, LSTM) і методів пояснюваного машинного навчання (XAI) для підвищення прозорості прогнозів і формування рекомендацій у режимі реального часу.

Перелік посилань

1. Brynjolfsson E., McAfee A. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W.W. Norton & Company, 2014. 306 p. URL: <https://wwnorton.com/books/the-second-machine-age/>
2. Goldfarb A., Tucker C. Digital Economics // *Journal of Economic Literature*. 2019. Vol. 57, No. 1. P. 3-43. DOI: 10.1257/jel.20171452
3. Montgomery D. C., Peck E. A., Vining G. G. *Introduction to Linear Regression Analysis*. 6th ed. Hoboken: Wiley, 2021. 704 p. URL: <https://www.wiley.com/en-us/Introduction+to+Linear+Regression+Analysis%2C+6th+Edition-p-9781119578727>
4. Jain A. K. Data Clustering: 50 Years Beyond K-Means // *Pattern Recognition Letters*. 2010. Vol. 31, No. 8. P. 651-666. DOI: 10.1016/j.patrec.2009.09.011
5. Ester M., Kriegel H.-P., Sander J., Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise // *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96)*. Portland, 1996. P. 226-231. URL: <https://file.biolab.si/papers/1996-DBSCAN-KDD.pdf>
6. Галузов С. Ю., Бондарчук А. П., Бажан Т. О., Корецька В. О. Застосування методів Data Science для прогнозування попиту в ритейлі // *Телекомунікаційні та інформаційні технології*. 2023. № 3. С. 56-62. DOI: 10.31673/2412-4338.2023.035965.
7. Бажан Т. О. Порівняльний аналіз методів машинного навчання для побудови прогнозів // *Сучасний захист інформації*. 2024. Т. 4 (60). С. 125-130. DOI: 10.31673/2409-7292.2024.040013.
8. Бажан Т. О. Інструменти очищення даних для прогнозування інвестицій за допомогою машинного навчання // *Інновації: Збірник наукових праць Державного університету інформаційно-комунікаційних технологій*. Київ: ДУІКТ, 2024. С. 45-50. URL: https://duikt.edu.ua/uploads/p_2779_56719466.pdf.

Надійшла 17.10.2025