

**Ганна Вайганг**

кандидат технічних наук, доцент

Національний університет біоресурсів і природокористування України, Київ, Україна

0000-0002-2082-2322

[weigang.ganna@nubip.edu.ua](mailto:weigang.ganna@nubip.edu.ua)

**Юрій Науринський**

Аспірант

Національний університет біоресурсів і природокористування України, Київ, Україна

0009-0004-6416-8635

[yu.naurynskiy@nubip.edu.ua](mailto:yu.naurynskiy@nubip.edu.ua)

## ТЕОРЕТИЧНА МОДЕЛЬ МЕНТАРИВНЕВОГО ВИБОРУ АЛГОРИТМІВ КЛАСТЕРИЗАЦІЇ ЗА ОПИСОВИМИ ХАРАКТЕРИСТИКАМИ НАБОРУ ДАНИХ ДЛЯ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

**Анотація.** Розглянуто теоретичні засади побудови ментарівневої моделі вибору алгоритмів кластеризації за описовими характеристиками набору даних для систем підтримки прийняття рішень. Запропоновано концептуальну схему, що забезпечує формалізований зв'язок між властивостями даних та алгоритмічними методами їх аналізу. Розроблена модель спрямована на підвищення адаптивності й пояснюваності аналітичних процедур, а також на покращення точності розпізнавання структур у даних і зменшення кількості помилкових об'єднань при кластеризації. Це створює підґрунтя для розвитку самонавчальних систем, здатних автоматично обирати оптимальні методи обробки даних залежно від їхньої структури та контексту застосування.

**Ключові слова:** системи підтримки прийняття рішень; алгоритм кластеризації; великі обсяги даних

### 1. ВСТУП

Сучасні системи підтримки прийняття рішень (СППР) працюють у середовищі великих, динамічних та неоднорідних даних. Їх ефективність значною мірою залежить від правильного вибору алгоритмів обробки та аналізу інформації, що визначають швидкість, точність і надійність отриманих результатів. Одним із ключових методів аналітичного етапу є кластеризація, яка дає змогу узагальнювати дані, виявляти приховані структури й формувати основи для подальших управлінських рішень.

Проблема полягає в тому, що різні алгоритми кластеризації демонструють неоднакову ефективність залежно від характеристик даних – їх розмірності, щільності, рівня шуму, топології та метрики подібності. Традиційно вибір алгоритму здійснюється вручну, що не відповідає вимогам сучасних СППР до автономності, адаптивності та масштабованості. Потребує розроблення теоретична модель, яка б дозволила формалізувати процес вибору алгоритмів кластеризації на основі описових характеристик набору даних і забезпечити адаптивну поведінку аналітичних підсистем.

**Мета публікації.** Теоретичне обґрунтування моделі ментарівневого вибору алгоритмів кластеризації за описовими характеристиками набору даних як інтелектуального компонента систем підтримки прийняття рішень, що забезпечує формальну відповідність між властивостями даних і методами їх аналітичної обробки, застосування.

### 2. ТЕОРЕТИЧНІ ОСНОВИ

Теоретичні основи побудови ментарівневої моделі спираються на уявлення про структуру даних як на сукупність параметрів, що можуть бути формалізовані та використані для прийняття рішень щодо вибору методів їх аналізу. Для того щоб алгоритм кластеризації був ефективним, він має відповідати властивостям конкретного набору даних, а отже, необхідно встановити зв'язок між характеристиками даних та

алгоритмічними підходами, здатними забезпечити стабільні результати. У цьому контексті важливим є опис простору дескрипторів, який дозволяє узагальнювати множину значущих параметрів і переводити задачу підбору алгоритму у формалізований процес зі зрозумілими критеріями відповідності.

Набір даних  $D$  характеризується множиною параметрів:

$$\delta(D) = \{d_1, d_2, \dots, d_n\} \quad (1)$$

де  $d_i$  відображають статистичні, структурні та інформаційні властивості даних: кількість об'єктів, розмірність, середню щільність, рівень шуму, кореляційні зв'язки, форму розподілу тощо [1].

Ці параметри утворюють простір дескрипторів  $\Delta$ , який узагальнює властивості наборів даних і дає можливість здійснювати метааналіз для вибору алгоритмів кластеризації.

Алгоритмічний простір визначимо як:

$$A = \{a_1, a_2, \dots, a_m\} \quad (2)$$

де кожен алгоритм  $a_j$  має свої припущення, обчислювальну складність і набір параметрів  $\theta(a_j)$ . Вибір алгоритму залежить від того, наскільки його властивості узгоджуються з дескрипторами даних. Зокрема, алгоритми метричного типу (наприклад, k-means) є ефективними для компактних і майже сферичних кластерів, тоді як щільнісні (DBSCAN, OPTICS) – для розріджених і шумових структур [2].

Функціональне відображення між простором дескрипторів і простором алгоритмів має вигляд:

$$f : \Delta \rightarrow A \quad (3)$$

де  $f$  визначає механізм вибору алгоритму кластеризації для кожного класу описових характеристик даних [3].

Задачу вибору можна подати як мінімізацію функції втрат:

$$L(f, D) = \ell(\text{результат кластеризації на } D \mid f(\delta(D))) \quad (4)$$

де  $\ell(\cdot)$  описує відхилення між структурою знайдених кластерів і бажаними властивостями аналітичного результату в межах СППР [4].

Такий підхід переводить задачу підбору алгоритму у метарівневий простір оптимізації, де об'єктом є не параметри кластеризації, а сама процедура вибору методу відповідно до властивостей даних.

Метарівнева модель функціонує як автономний аналітичний модуль, що передусь безпосередній кластеризації. Її робота складається з трьох етапів:

1. Аналіз дескрипторів набору даних та побудова вектору характеристик.
2. Визначення класу даних і відображення його у простір алгоритмів кластеризації за допомогою функції  $f$ .
3. Формування рекомендацій або автоматичний вибір алгоритму з відповідними гіперпараметрами [5].

Таким чином забезпечується самоналаштуваність аналітичного рівня СППР, скорочується час прийняття рішень і підвищується стабільність результатів у динамічних середовищах. Метарівнева модель також створює основу для подальшого розвитку адаптивних систем, у яких відбір алгоритмів стає частиною процесу навчання системи. На практиці це означає, що система здатна накопичувати досвід попередніх

запусків: якщо певний алгоритм демонструє стабільно кращі показники для схожих наборів, він буде рекомендований автоматично.

### **3. МЕТОДИ ДОСЛІДЖЕННЯ**

Дослідження базується на метарівневому аналітичному підході, який поєднує методи системного аналізу, узагальнення наукових джерел та теоретичного моделювання. На першому етапі було проведено огляд публікацій і порівняльних досліджень, присвячених алгоритмам кластеризації, AutoML-підходам і мета-навчанню [1]–[3]. Це дозволило виявити закономірності між властивостями даних і результативністю окремих алгоритмів, що стало основою для побудови метарівневої моделі вибору.

На другому етапі сформульовано концепцію тривірневої моделі, яка описує взаємозв'язок між характеристиками набору даних, множиною алгоритмів кластеризації та процесом прийняття рішення у СППР [4], [5]. Такий підхід забезпечує теоретичне підґрунтя для подальшого розвитку адаптивних і самонавчальних систем підтримки прийняття рішень.

### **4. РЕЗУЛЬТАТИ ТА ОБГОВОРЕННЯ**

У результаті дослідження було сформовано концептуальну основу метарівневої моделі вибору алгоритмів кластеризації для систем підтримки прийняття рішень. Запропонована модель розглядається як проміжна ланка між рівнем даних та аналітичним рівнем системи, що забезпечує автоматизований вибір методу кластеризації залежно від характеристик набору даних. Вона передбачає наявність функціонального зв'язку між описовими параметрами даних, такими як кількість об'єктів, розмірність, щільність, рівень шуму, структура кореляцій, та класом алгоритмів, який найкраще відповідає цим властивостям. Це дає змогу системі здійснювати вибір алгоритму без участі користувача, що зменшує вплив людського фактора та підвищує адаптивність і відтворюваність аналітичних процесів.

Наприклад, у наборах даних з високою розрідженістю (транспортні потоки, журнали подій) модель обирає HDBSCAN, тоді як для компактних і збалансованих ознак (медичні таблиці або результати соціальних опитувань) доцільними є GMM або k-means.

Модель інтегрується в архітектуру систем підтримки прийняття рішень як надбудова аналітичного модуля, яка виконує попередню обробку вхідних даних, аналізує їх дескриптори та визначає відповідний алгоритм кластеризації для подальшої обробки. Такий підхід дозволяє зменшити кількість обчислювальних експериментів, пов'язаних із пошуком оптимального методу, а також скорочує час отримання результатів, що є критично важливим у системах реального часу. Крім того, модель забезпечує структурованість і логічну прозорість процесу вибору, що сприяє підвищенню пояснюваності аналітичних рішень у межах СППР.

Проведений аналіз показує, що метарівнева модель створює основу для побудови самонавчальних аналітичних систем, у яких результати попередніх обчислень можуть бути використані як база знань для подальших виборів. Це відкриває можливість накопичення метаінформації про ефективність алгоритмів у різних сценаріях, що у перспективі дозволить формувати узагальнену базу рекомендацій. Такі системи можуть еволюціонувати з часом, адаптуючи власну поведінку до змін у структурі даних та зовнішніх умовах.

Перспективи подальшого розвитку дослідження полягають у формуванні багатокритеріальної моделі, яка враховуватиме не лише структурні властивості даних, але й показники продуктивності, такі як час виконання, споживання ресурсів, стійкість

і точність. Це дозволить реалізувати комплексну систему вибору, у якій кожен алгоритм оцінюється за вектором критеріїв, що забезпечить оптимізацію не лише якості кластеризації, а й ефективності всієї системи підтримки прийняття рішень. У подальшому передбачається розроблення адаптивної метамоделі навчання, здатної оновлювати правила вибору алгоритмів на основі накопичених знань, що наблизить СППР до концепції повністю інтелектуальної, самонавчальної системи аналітичного прийняття рішень.

### **ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ**

У ході дослідження розроблено теоретичну модель метарівневого вибору алгоритмів кластеризації за описовими характеристиками набору даних, яка розглядається як ключовий компонент адаптивних систем підтримки прийняття рішень. Запропонований підхід дозволяє формалізувати процес вибору алгоритму, зменшити залежність від експертних оцінок та забезпечити узгодженість між властивостями даних і методами їх аналізу. Модель зменшує кількість ручних налаштувань при виборі алгоритму і скорочує час отримання коректного результату кластеризації за рахунок автоматизованого аналізу властивостей даних. Перспективами подальших досліджень є розроблення практичних методів реалізації метарівневої моделі у вигляді програмних компонентів, інтегрованих із існуючими системами аналізу даних, а також розширення її можливостей шляхом включення критеріїв продуктивності, масштабованості та самооновлення на основі накопиченого досвіду роботи системи.

### **ПОСИЛАННЯ**

- [1] Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2), 165–193. <https://doi.org/10.1007/s40745-015-0040-1>
- [2] Garouani, M., Ahmad, A., Bouneffa, M., Hamlich, M., Bourguin, G., & Lewandowski, A. (2022). Using meta-learning for automated algorithms selection and configuration: an experimental framework for industrial big data. *Journal of Big Data*, 9(1). <https://doi.org/10.1186/s40537-022-00612-4>
- [3] Poulakis, Y., Doulkeridis, C., & Kyriazis, D. (2024). A survey on AutoML Methods and Systems for Clustering. *ACM Transactions on Knowledge Discovery From Data*, 18(5), 1–30. <https://doi.org/10.1145/3643564>
- [4] Barbudo, R., Ventura, S., & Romero, J. R. (2023). Eight years of AutoML: categorisation, review and trends. *Knowledge and Information Systems*, 65(12), 5097–5149. <https://doi.org/10.1007/s10115-023-01935-1>
- [5] Rivolli, A., Garcia, L. P., Soares, C., Vanschoren, J., & De Carvalho, A. C. (2022). Meta-features for meta-learning. *Knowledge-Based Systems*, 240, 108101. <https://doi.org/10.1016/j.knosys.2021.108101>

MINISTRY OF EDUCATION  
AND SCIENCE OF UKRAINE

NATIONAL UNIVERSITY  
OF LIFE AND ENVIRONMENTAL  
SCIENCES OF UKRAINE

FACULTY OF INFORMATION  
TECHNOLOGY

МІНІСТЕРСТВО ОСВІТИ  
І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
БІОРЕСУРСІВ І  
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ  
ТЕХНОЛОГІЙ

## PROCEEDINGS

XIII International scientific  
and practical conference

**GLOBAL AND  
REGIONAL PROBLEMS OF  
INFORMATIZATION IN  
SOCIETY AND  
NATURE USING  
'2025**

13-14 November 2025

Kyiv, NULES of Ukraine

Kyiv 2025

## МАТЕРІАЛИ

XIII Міжнародної науково-  
практичної конференції

**ГЛОБАЛЬНІ ТА  
РЕГІОНАЛЬНІ ПРОБЛЕМИ  
ІНФОРМАТИЗАЦІЇ В  
СУСПІЛЬСТВІ І  
ПРИРОДОКОРИСТУВАННІ  
'2025**

13-14 листопада 2025 року

Київ, НУБіП України

Київ 2025

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ  
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ  
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

## **МАТЕРІАЛИ**

ХІІІ Міжнародної науково-практичної конференції

# **ГЛОБАЛЬНІ ТА РЕГІОНАЛЬНІ ПРОБЛЕМИ ІНФОРМАТИЗАЦІЇ В СУСПІЛЬСТВІ І ПРИРОДОКОРИСТУВАННІ '2025**

13-14 листопада 2025 року

Київ, НУБіП України

Київ 2025

УДК 004

Рекомендовано до друку вченою радою факультету інформаційних технологій Національного університету біоресурсів і природокористування України (протокол № 4 від 18.12.2025).

Укладач: д.т.н., доцент Шкарупило В.В.

Збірник матеріалів XIII Міжнародної науково-практичної конференції "Глобальні та регіональні проблеми інформатизації в суспільстві і природокористуванні '2025", 13–14 листопада 2025 року, НУБіП України, Київ. – К.: НУБіП України, 2025. – 206 с.

Відповідальність за зміст публікацій несуть автори.

© Національний університет біоресурсів  
і природокористування України, 2025