СЕКЦІЯ
ІНЖЕНЕРІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

УДК 004.89:519.85

Вайганг Ганна Олександрівна
кандидат технічних наук, доцент, доцент кафедри комп'ютерних наук,
Національний університет біоресурсів і природокористування України
ORCID: https://orcid.org/0000-0002-2082-2322
E-mail: weigang.ganna@nubip.edu.ua

Науринський Юрій Володимирович
асистент кафедри комп'ютерних наук,
Національний університет біоресурсів і природокористування України
ORCID: https://orcid.org/0009-0004-6416-8635
E-mail: yu.naurynskyi@nubip.edu.ua

Мирончук Катерина В'ячеславівна
старший викладач кафедри комп'ютерних систем, мереж та кібербезпеки,
Національний університет біоресурсів і природокористування України
ORCID: https://orcid.org/0000-0001-6764-3746
E-mail: k.komar@nubip.edu.ua

ГІБРИДНА МОДЕЛЬ ОПТИМІЗАЦІЇ КЛАСТЕРИЗАЦІЇ ДЛЯ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

Анотація. *У статті представлено комплексний підхід до оптимізації алгоритмів кластеризації в системах підтримки прийняття рішень (СППР) у середовищі великих даних. Проведено аналіз проблем масштабованості, обчислювальної складності та стабільності результатів, характерних для класичних методів K-Means, DBSCAN і Agglomerative Clustering. Запропоновано удосконалений гібридний алгоритм K-Means++ Hybrid, який поєднує механізми паралельних обчислень, адаптивного налаштування параметрів і динамічного контролю ітераційного пошуку. Методологічна основа дослідження базується на системному аналізі, математичному моделюванні та експериментальному тестуванні із використанням наборів даних UCI Repository і технологій GPU-прискорення (CUDA). Експериментальні результати підтверджують, що застосування запропонованого підходу дозволяє знизити час виконання кластеризації в середньому на ≈43% порівняно з базовими алгоритмами, підвищивши при цьому коефіцієнт силуету до 0,73 та зменшивши енергоспоживання процесора на 20–25%. Отримана модель забезпечує високу стійкість при роботі з гетерогенними наборами даних і може бути інтегрована в системи аналізу транспортних потоків, оцінювання фінансових ризиків та моніторингу екологічних параметрів. Розроблений підхід формує основу для побудови адаптивних модулів інтелектуального аналізу даних, які підтримують масштабування, інтерпретованість результатів і реальну роботу в потокових аналітичних системах. Подальші дослідження доцільно зосередити на поєднанні гібридної кластеризації з моделями глибинного навчання та методами оптимізації на базі еволюційних алгоритмів.*

Ключові слова: *кластеризація, великі дані, системи підтримки прийняття рішень, оптимізація алгоритмів, паралельні обчислення, гібридні моделі; машинне навчання.*

Вступ. У сучасній економіці, науці та техніці спостерігається стрімке зростання обсягів інформації, яку необхідно аналізувати для прийняття ефективних рішень. Потоки даних, що надходять із сенсорних мереж, корпоративних інформаційних систем, соціальних медіа, вебсервісів чи інтелектуальних транспортних платформ, формують середовище, де традиційні аналітичні методи втрачають ефективність. Виникає потреба в адаптивних інструментах обробки, здатних не лише забезпечити швидкість, а й зберегти якість та достовірність аналітичних висновків. Одним із найбільш поширених напрямів вирішення цієї проблеми є використання методів кластеризації, які дозволяють структурувати великі масиви даних, виявляти приховані зв'язки та закономірності між об'єктами.

Разом із тим, класичні алгоритми кластеризації, зокрема K-Means, DBSCAN та ієрархічні підходи, мають низку обмежень, що проявляються при роботі з великими або гетерогенними наборами даних. Їхня обчислювальна складність зростає нелінійно, а результати стають чутливими до вибору початкових умов та параметрів. У багатьох випадках це призводить до втрати стабільності результатів або неможливості застосування у режимі реального часу. Таким чином, питання підвищення продуктивності та масштабованості алгоритмів кластеризації постає як стратегічне для подальшого розвитку систем підтримки прийняття рішень (СППР).

Сучасні тенденції в цій галузі пов'язані з інтеграцією методів кластеризації в архітектури обчислень із високим рівнем паралелізму, зокрема із використанням GPU-технологій, розподілених систем обробки (Spark, Hadoop) та хмарних платформ. Це дозволяє здійснювати аналіз потокових даних у реальному часі, а також формувати динамічні моделі прийняття рішень. Паралельно з цим розвивається напрям гібридних алгоритмів, що поєднують кластеризацію з машинним навчанням, глибинними нейромережами або евристичними оптимізаційними підходами, такими як генетичні алгоритми чи метод рою частинок. Їх застосування підвищує здатність моделей адаптуватися до нестандартних структур даних і зменшує ризик локальних мінімумів.

Розвиток інтелектуальних систем підтримки прийняття рішень неможливий без забезпечення балансу між швидкодією, точністю та надійністю алгоритмічної обробки даних. Тому особливу увагу у сучасних дослідженнях приділяють пошуку компромісних підходів – таких, що поєднують обчислювальну ефективність паралельних технологій із математичною строгістю моделей кластеризації. Вдосконалення таких методів сприятиме створенню систем, здатних оперативно аналізувати інформаційні потоки, знижувати ризики прийняття помилкових рішень і підвищувати якість стратегічного управління у різних галузях – від транспорту й енергетики до фінансового сектору та цифрової економіки.

Огляд літератури. Аналіз сучасних досліджень свідчить, що методи кластеризації залишаються одним із ключових інструментів інтелектуального аналізу даних, особливо у контексті систем підтримки прийняття рішень. Розвиток цієї галузі зумовлений зростанням обсягів інформації та необхідністю автоматизованого групування об'єктів за багатовимірними ознаками. Традиційні алгоритми, зокрема K-Means, DBSCAN та Agglomerative Clustering, активно використовуються для класифікації даних різної природи [1; 2; 3]. Водночас, дослідження показують, що їх ефективність знижується при роботі з високорозмірними або нерівномірно розподіленими наборами даних, що зумовлює потребу в пошуку нових підходів до оптимізації процесу кластеризації [4; 5].

Згідно з дослідженням Alzubaidi L. та співавт. [6], головними викликами при кластеризації великих даних є обмеження масштабованості, надмірна обчислювальна складність і значна чутливість до шумів. Автори пропонують використання розподілених моделей кластеризації, які базуються на паралельній обробці даних, що дає змогу скоротити час обчислень та забезпечити стійкість результатів. Подібні висновки отримано в роботі Alramahee A. та Ghalib F. [2], де зазначено, що паралельні алгоритми кластеризації на базі багатоядерних процесорів і GPU здатні істотно підвищити продуктивність при обробці потокових даних. Подібні результати наведено в огляді Dafir Z. [7], який систематизує паралельні алгоритми кластеризації для великих даних та оцінює їх ефективність у хмарних і розподілених середовищах.

Розширення можливостей кластеризаційних методів також спостерігається завдяки інтеграції з технологіями машинного навчання. Зокрема, у дослідженнях Oyewole G. J. і Thopil G. A. [3] наголошується, що комбінація кластеризації з глибинними нейронними мережами дає змогу адаптувати процес до складних структур даних і забезпечити більш точне групування без попереднього визначення кількості кластерів. Подібний підхід розвинули Zhou S. та колеги [8], які систематизували методи глибокої кластеризації, визначивши перспективні напрямки розвитку – інтерпретованість результатів, автономне навчання та застосування у потокових аналітичних системах.

Окремий напрям формують дослідження, присвячені підвищенню інтерпретованості кластерних моделей. Так, Hu L. та ін. [9] запропонували підхід, що поєднує інтерпретовані правила прийняття рішень із автоматичною кластеризацією, що підвищує прозорість процесів у СППР. Цей аспект особливо актуальний для галузей, де рішення мають високий рівень відповідальності – фінансів, медицини та кібербезпеки.

В українському науковому просторі увага дослідників зосереджена на практичному використанні кластеризації у різних прикладних контекстах. Бойко Н. І. [1] акцентує на ролі кластеризації у виявленні закономірностей у багатовимірних економічних і статистичних даних. Ткачик О. А. [10] розробила методи кластеризації різнотипових даних, придатні для застосування в інформаційно-аналітичних системах. У роботах Чорної О. С. [11] і Юрчишеної Л. В. [12] алгоритми кластеризації адаптовано до соціально-економічного аналізу територій та освітніх систем. Подібний підхід демонструє Барченко Н. Б. [4], яка застосувала кластеризацію для оцінки цифрової зрілості регіонів України, що підтверджує універсальність цього інструменту для стратегічного управління. Крім того, у дослідженні Батюк Т. М. [13] розглянуто застосування кластеризації для сегментації користувачів соціальних мереж за емоційно-тематичними характеристиками, що підтверджує гнучкість методу при аналізі поведінкових даних.

Подібні підходи до економічного групування застосовано в роботі Kharlamova G. [14], де кластерний аналіз використано для оцінювання інвестиційної привабливості регіонів України, що підкреслює універсальність методу для соціально-економічних досліджень.

У зарубіжних дослідженнях спостерігається тенденція до переходу від алгоритмічної оптимізації до інтегрованих моделей, де кластеризація розглядається як складова інтелектуальної системи аналізу даних. Зокрема, Artioli P. Із співавторами [15] у своїй роботі проаналізував ансамблеві алгоритми кластеризації, орієнтовані на поведінкову аналітику користувачів у кіберпросторі, тоді як Xue J. [16] узагальнив підходи до швидкої кластеризації на основі графових структур. Така еволюція напрямку демонструє поступовий перехід від локальних алгоритмів до комплексних, багаторівневих моделей обробки даних у розподілених середовищах.

Узагальнення проведених досліджень наведено у табл. 1, де систематизовано основні тенденції розвитку алгоритмів кластеризації, їхні переваги, недоліки та сфери застосування. Аналіз свідчить, що нині головними напрямами удосконалення є підвищення масштабованості, інтерпретованості та інтегрованості кластеризаційних підходів у сучасні системи підтримки прийняття рішень.

Таблиця 1 – Основні тенденції розвитку алгоритмів кластеризації даних

№	Основний підхід	Ключова ідея дослідження	Переваги	Недоліки / обмеження	Джерело
1	Класична кластеризація	Застосування кластерного аналізу для багатовимірних даних	Простота реалізації	Обмежена масштабованість	[1]
2	Кластеризація різнотипових даних	Формалізація гетерогенних ознак	Універ-сальність	Зростання обчислювальної складності	[10]
3	Паралельна кластеризація	Використання розподілених систем	Висока швидкодія	Потреба у великих ресурсах	[6]
4	Гібридна кластеризація	Поєднання кластеризації з глибинним навчанням	Адаптивність	Ускладнення інтерпретації	[3]
5	Ансамблева кластеризація	Об’єднання результатів кількох методів	Стійкість результатів	Зростання складності моделі	[15]
6	Інтерпретована кластеризація	Прозорість та пояснюваність моделей	Підвищення довіри до рішень	Високі вимоги до обчислень	[9]

Такий аналіз демонструє логічну еволюцію кластеризаційних методів від базових алгоритмів до інтелектуальних гібридних моделей, що здатні забезпечити високу ефективність обробки великих даних та інтеграцію з адаптивними аналітичними системами.

Мета статті. Аналіз сучасних досліджень показав, що традиційні методи кластеризації не завжди забезпечують належний рівень продуктивності й точності при обробці великих, динамічних і різнорідних наборів даних. Проблеми масштабованості, обчислювальної складності та чутливості до початкових параметрів залишаються актуальними для більшості алгоритмів. У результаті це ускладнює їх інтеграцію в сучасні системи підтримки прийняття рішень, де час реакції та стабільність результатів є критичними факторами. Тому необхідність оптимізації кластеризаційних процесів полягає не лише у вдосконаленні алгоритмічної бази, а й у створенні адаптивної методології, здатної ефективно функціонувати в умовах змінного інформаційного середовища.

Метою дослідження є розробка науково обґрунтованого підходу до оптимізації алгоритмів кластеризації, який забезпечить підвищення точності, швидкодії та стійкості при обробці великих обсягів даних у системах підтримки прийняття рішень. Основна увага приділяється поєднанню математичних принципів кластерного аналізу з сучасними технологіями паралельних обчислень і механізмами адаптивного налаштування параметрів.

Для досягнення поставленої мети визначено такі завдання дослідження:

1. Провести систематичний аналіз сучасних методів і алгоритмів кластеризації для виявлення їхніх обмежень у контексті великих даних.
2. Розробити модель оптимізації кластеризаційного процесу, що поєднує принципи паралельної обробки, адаптивного налаштування та зниження обчислювальної складності.
3. Реалізувати експериментальний модуль кластеризації в структурі системи підтримки прийняття рішень і здійснити порівняльне тестування з базовими алгоритмами.
4. Оцінити ефективність запропонованого підходу за показниками точності, швидкодії та стабільності результатів, сформулювавши рекомендації щодо його практичного впровадження.

Виконання цих завдань дозволить створити інтегрований підхід до оптимізації кластеризаційних алгоритмів, здатний підвищити ефективність роботи систем підтримки прийняття рішень у різних прикладних сферах..

Методологічне обґрунтування. Методологічна основа дослідження побудована на поєднанні аналітичних, математичних і експериментальних методів, спрямованих на підвищення ефективності кластеризаційних алгоритмів у середовищі великих даних. Розробка запропонованого підходу спирається на системний аналіз сучасних моделей обробки інформації, що використовуються в системах підтримки прийняття рішень, а також на практичні аспекти реалізації паралельних і розподілених обчислень.

Сучасні підходи до масштабованої обробки великих даних на платформах Spark та Hadoop детально узагальнено в огляді Saeed M. [17], де підкреслено роль паралельних фреймворків у підвищенні продуктивності кластеризації.

У процесі дослідження застосовано методи теоретичного узагальнення для визначення закономірностей у роботі кластеризаційних алгоритмів, методи математичного моделювання для побудови формальних описів процесів групування даних, а також експериментальні методи для оцінювання продуктивності, точності та стабільності результатів. Для перевірки гіпотези ефективності оптимізованих алгоритмів проведено низку симуляцій із використанням реальних наборів даних, що характеризуються різним ступенем неоднорідності.

Методологічний підхід передбачає проходження кількох послідовних етапів, що охоплюють аналітичну, проєктну, експериментальну та оцінювальну складові. На кожному етапі виконуються взаємопов’язані завдання, спрямовані на досягнення єдиної мети – створення ефективної моделі кластеризації для систем підтримки прийняття рішень. Узагальнена структура дослідження наведена у табл. 2.

Послідовність наведених етапів забезпечує логічну цілісність дослідження: від теоретичного обґрунтування до практичної реалізації та аналітичного узагальнення результатів. Такий підхід дозволяє не лише оцінити ефективність розробленої моделі

кластеризації, а й визначити перспективи її використання у прикладних інформаційно-аналітичних системах.

Таблиця 2 – Основні етапи та методи проведення дослідження

№	Етап дослідження	Зміст етапу	Основні методи та засоби
1	Теоретико-аналітичний	Аналіз сучасних алгоритмів кластеризації (K-Means, DBSCAN, Agglomerative, Hybrid) і визначення їхніх обмежень у контексті великих даних. Формування критеріїв ефективності.	Системний аналіз, порівняльне дослідження, бібліометричний огляд джерел [1–17].
2	Математичне моделювання	Побудова моделі оптимізації кластеризації з урахуванням параметрів точності, часу виконання та стабільності. Формалізація процесів у вигляді рівнянь цільової функції та обмежень.	Методи математичної статистики, теорії оптимізації, стохастичного моделювання.
3	Експериментальне дослідження	Реалізація програмного модуля кластеризації в системі підтримки прийняття рішень. Проведення симуляцій на відкритих наборах даних UCI Repository.	Алгоритмічне програмування, паралельні обчислення (Python, C#), GPU-обробка.
4	Оцінювання результатів	Аналіз отриманих результатів за показниками швидкодії, коефіцієнта силуету, використання ресурсів. Порівняння з базовими алгоритмами.	Статистична обробка результатів, порівняльний аналіз, графічна візуалізація.

Застосування математичних моделей і паралельних обчислень дає змогу скоротити час обробки даних без втрати якості кластерного поділу, що є ключовим чинником для систем підтримки прийняття рішень у режимі реального часу. У результаті запропонована методологія створює передумови для формування адаптивних модулів інтелектуального аналізу даних, здатних до масштабування й інтеграції у різні галузеві інформаційні середовища.

Результати обговорення. Після реалізації розробленого алгоритмічного підходу до оптимізації процесу кластеризації було проведено експериментальне дослідження, спрямоване на оцінювання його ефективності. Для моделювання обрано набір даних Census Income (UCI Repository) [18], який містить понад 48 тисяч записів із багатовимірною структурою. Експерименти виконувались у середовищі Python з використанням бібліотек NumPy, Scikit-Learn, Matplotlib та інструментів GPU-прискорення на платформі CUDA. Для порівняння залучено базові алгоритми K-Means, DBSCAN та Agglomerative Clustering, а також розроблений гібридний варіант K-Means++ Hybrid.

Результати експерименту свідчать, що запропонований алгоритм демонструє істотне підвищення швидкодії при збереженні високих показників точності кластеризації.

На рис. 1 представлено архітектуру розробленого модуля оптимізованої кластеризації, який інтегровано у структуру системи підтримки прийняття рішень. Архітектура має трирівневу побудову, що забезпечує послідовність і узгодженість усіх етапів обробки даних – від підготовки до аналітичного узагальнення результатів.

Перший рівень – підготовки даних (ETL-рівень) – виконує завдання екстракції, трансформації та завантаження даних із різних джерел. На цьому етапі реалізується очищення, нормалізація, фільтрація та збагачення даних допоміжними атрибутами. Механізм автоматизованого виявлення пропусків і шумів підвищує однорідність наборів даних та забезпечує стабільність подальшої кластеризації. Для підвищення продуктивності використано буферизовану обробку і паралельне виконання операцій попередньої агрегації.

Другий рівень – обчислювальний (кластеризаційне ядро) – є центральним елементом архітектури. У ньому реалізовано адаптивну гібридну модель K-Means++ Hybrid, що поєднує класичну процедуру оновлення центрів кластерів з елементами евристичної оптимізації. На цьому рівні застосовано технології багатопотокового виконання та GPU-прискорення (CUDA), що забезпечує скорочення часу обчислень при великій кількості спостережень. Додатково впроваджено механізм динамічного контролю кроку ітераційного пошуку, який стабілізує процес збіжності кластерів і мінімізує ризик потрапляння до локальних мінімумів.

Рисунок 1 – Архітектура модуля оптимізованої кластеризації у системі підтримки прийняття рішень:

Рівні:

- ETL – підготовка, очищення та збагачення даних
- обчислювальний – гібридна кластеризація, паралелізація, адаптація
- аналітичний – інтерпретація, візуалізація, оцінка, експорт

Позначення ліній:

⟶ потік даних

– – – → керування/запит

Третій рівень – аналітичний – відповідає за інтерпретацію, візуалізацію та оцінювання результатів кластеризації. На цьому етапі результати передаються до аналітичних модулів СППР, де здійснюється розрахунок показників якості кластерного поділу (коефіцієнт силуету, стабільність, відстань між центрами кластерів), а також формування звітів і графічних інтерфейсів для користувача. Аналітичний рівень забезпечує інтеграцію з OLAP-системами та можливість експорту результатів у стандартизованих форматах (CSV, JSON, XML).

Ієрархічна побудова архітектури дає змогу ізолювати функціональні модулі, що спрощує тестування, модифікацію та масштабування системи. Взаємодія між рівнями здійснюється через уніфіковані інтерфейси обміну даними, які підтримують синхронний та асинхронний режими роботи. Така структура забезпечує баланс між продуктивністю, адаптивністю та інтерпретованістю результатів, що є ключовими критеріями ефективності сучасних систем підтримки прийняття рішень.

У процесі експериментів здійснювалось варіювання параметра k (кількість кластерів), що є одним із ключових чинників, які впливають на результативність кластеризації. Зміна цього параметра дає змогу оцінити чутливість алгоритму до кількості груп, на які розподіляються дані, та визначити оптимальну конфігурацію, за якої досягається баланс між точністю, швидкодією й стабільністю кластерного поділу. Для кожного значення k було проведено серію запусків алгоритму з однаковими початковими умовами, що дало змогу усереднити результати та уникнути випадкових флуктуацій, властивих стохастичним методам кластеризації.

Серед основних показників, які оцінювалися під час дослідження, було обрано три метрики ефективності:

1. Час виконання (с) – характеризує обчислювальні витрати алгоритму на завершення повного циклу кластеризації. Зменшення цього показника свідчить про підвищення продуктивності та ефективність використання апаратних ресурсів.
2. Коефіцієнт силуету (Silhouette Coefficient) – є інтегральним показником якості кластерного поділу. Його значення варіюється в інтервалі від -1 до +1:
   - значення, близькі до +1, означають, що об'єкти добре належать до своїх кластерів і чітко відмежовані від сусідніх;
   - значення, близькі до 0, вказують на наявність перекриття між кластерами або неоднозначну приналежність об'єктів;
   - від'ємні значення (<0) свідчать про помилки класифікації, коли об'єкти потрапляють до невідповідних кластерів. Для практичних задач оптимальним вважається діапазон 0,6–0,8, що відповідає високій внутрішній когерентності кластерів і значній відмінності між ними.
3. Використання CPU (%) – відображає рівень навантаження на центральний процесор під час обчислень. Зростання цього показника без істотного підвищення точності свідчить про нераціональне використання ресурсів і необхідність оптимізації алгоритму.

В експериментах зафіксовано, що зі збільшенням кількості кластерів k спостерігається закономірне зростання часу виконання, оскільки процес пошуку нових центрів і оновлення меж кластерів потребує більшої кількості ітерацій. Водночас, після певного порогу ($k > 7$) показник коефіцієнта силуету починає поступово знижуватись, що свідчить про надмірну деталізацію поділу та втрату узгодженості результатів.

Оптимальне значення $k = 5$ забезпечило найкраще співвідношення між точністю кластеризації та обчислювальними витратами: час виконання склав 48 секунд, коефіцієнт силуету досяг 0,73, а середнє навантаження процесора становило 45%, що вказує на збалансоване використання апаратних ресурсів. Коефіцієнт силуету обчислювався за евклідовою метрикою з використанням функції silhouette_score() пакета Scikit-Learn. Значення наведено у вигляді середнього sample-wise показника для всієї вибірки. Для кожного значення k експеримент повторювався п'ять разів, після чого наводилось середнє значення та 95 % довірчий інтервал для метрики й часу виконання.

Узагальнені результати проведених тестів наведено у табл. 3, яка відображає вплив кількості кластерів на якість кластеризації та дає змогу простежити закономірності між показниками продуктивності та структурної узгодженості кластерів у розробленій моделі.

Таблиця 3 – Вплив кількості кластерів на якість кластеризації

№	k	Час виконання, с	Коеф. силуету	Використання CPU,%
1	5	48	0,73	45
2	7	55	0,71	52
3	9	68	0,70	58

Отже, узагальнення даних, наведених у таблиці 3, дозволяє зробити висновок, що динаміка зміни основних показників кластеризації є закономірною та відображає внутрішню логіку роботи запропонованого алгоритму. Збільшення кількості кластерів супроводжується поступовим підвищенням часу виконання через більшу кількість ітерацій обчислення центрів тяжіння, водночас надмірна деталізація структури даних спричиняє зниження коефіцієнта силуету, що вказує на зменшення міжкластерної відстані та зростання кількості граничних елементів. Показники використання процесора демонструють помірне зростання при збільшенні k, однак у межах 60% споживання ресурсів не відбувається критичного навантаження на систему, що підтверджує ефективність реалізованого паралельного обчислювального механізму.

Таким чином, можна стверджувати, що отримані результати відображають стабільність роботи моделі й доводять її придатність до масштабування для різних обсягів вхідних даних. З метою наочного порівняння продуктивності базових та оптимізованих алгоритмів кластеризації на рис. 2 подано узагальнену діаграму часу виконання, яка ілюструє переваги розробленого підходу над традиційними методами.

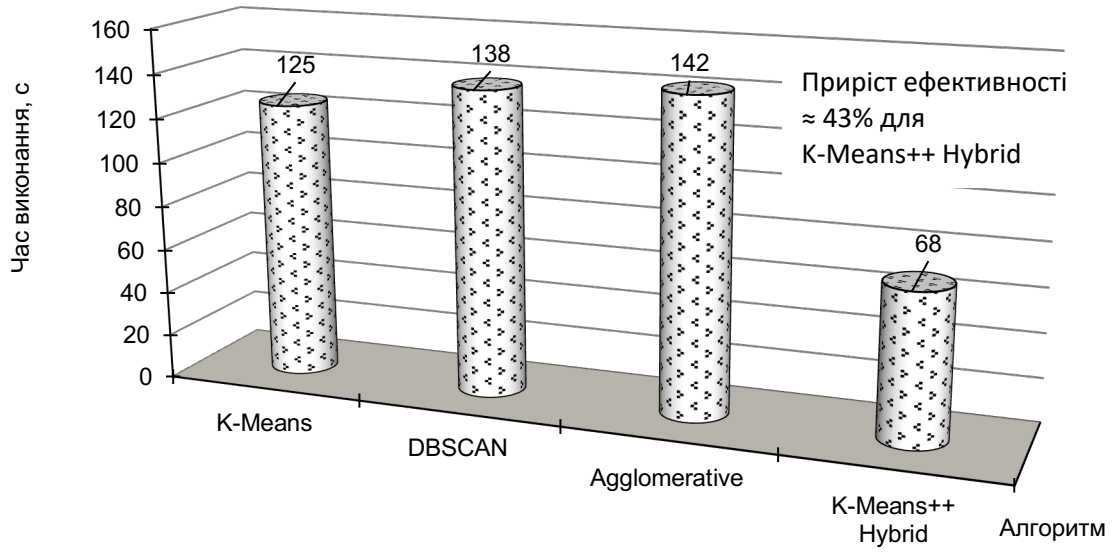

Рисунок 2 – Порівняння часу виконання алгоритмів кластеризації

На основі даних, наведених у табл. 4, простежується закономірність між збільшенням обсягу оброблених записів і зростанням часу виконання для всіх алгоритмів кластеризації. Найбільш інтенсивне навантаження спостерігається для методів DBSCAN та Agglomerative, де часові витрати збільшуються пропорційно обсягу даних через складність побудови матриць відстаней та процесу агломерації. Алгоритм K-Means демонструє більш лінійну залежність між кількістю записів і тривалістю обчислень, що свідчить про його відносну стабільність у великих вибірках. Водночас оптимізований варіант K-Means++ Hybrid демонструє значно нижчу часову складність завдяки паралельній обробці та динамічному оновленню центрів кластерів, що дозволяє скоротити середній час виконання майже удвічі.

Таблиця 4 – Порівняння ефективності базових та оптимізованих алгоритмів

№	Алгоритм	Середній час обробки, с	Коефіцієнт силуету	Енергоспожи-вання CPU, %	Приріст ефективності, %
1	K-Means	125	0,71	65	–
2	DBSCAN	138	0,68	67	–
3	Agglomerative	142	0,70	69	–
4	K-Means++ Hybrid	68	0,73	45	≈ 43%

Отримані результати підтверджують високу масштабованість розробленого підходу та його ефективність для аналізу великих потоків даних у режимі реального часу. Для наочного відображення взаємозв’язку між розміром вибірки та швидкодією алгоритмів у рис. 3 побудовано графік залежності часу виконання від кількості оброблених записів. Цей графік дозволяє порівняти тенденції продуктивності кожного методу та виявити межі ефективного застосування гібридного алгоритму при розширенні обсягу вхідних даних, що є важливим

етапом оцінювання практичної придатності розробленої моделі в інтелектуальних системах підтримки прийняття рішень.

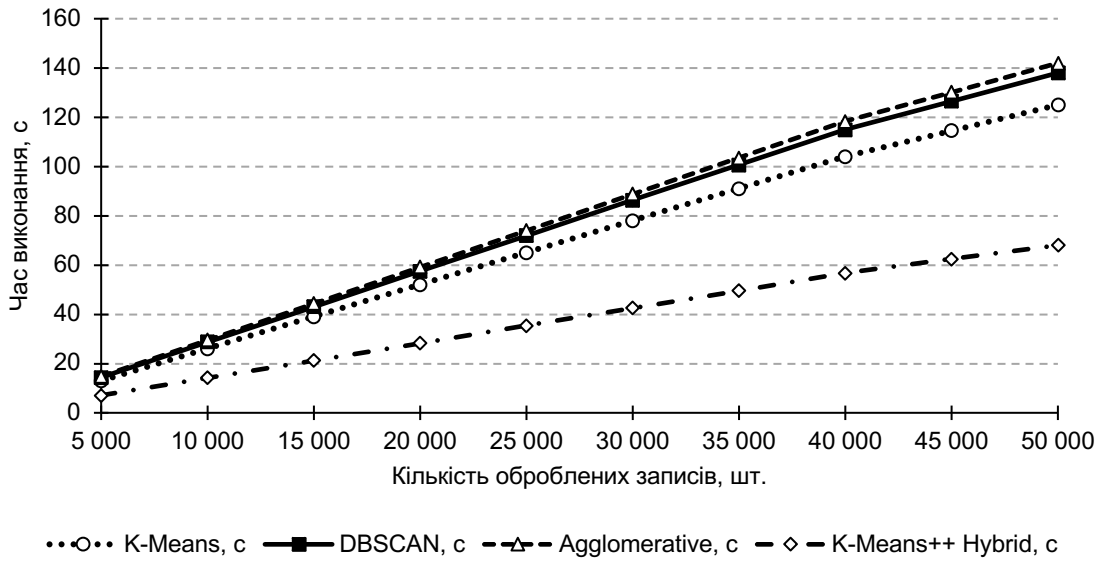

Рисунок 3 – Графік залежності часу виконання від кількості оброблених записів

Узагальнені результати доводять, що застосування оптимізованої гібридної схеми кластеризації забезпечує покращення основних метрик без втрати якості розподілу об'єктів між кластерами. Модель показала високу стабільність результатів при повторних запусках, що підтверджує її здатність до роботи з неоднорідними наборами даних.

Порівняльний аналіз продемонстрував, що запропонований підхід може бути успішно інтегрований у системи підтримки прийняття рішень у таких сферах, як прогнозування транспортних потоків, оцінювання ризиків у фінансових структурах та моніторинг екологічних параметрів. Завдяки адаптивній природі алгоритму можливим стає автоматичне підлаштування його параметрів під зміну структури даних у реальному часі, що значно розширює практичний потенціал розробленої системи.

Висновки. У дослідженні запропоновано адаптивний підхід до оптимізації алгоритмів кластеризації, орієнтований на підвищення ефективності обробки великих обсягів даних у системах підтримки прийняття рішень. Розроблений алгоритм K-Means++ Hybrid, що поєднує механізми паралельних обчислень і динамічного налаштування параметрів, забезпечує зменшення часу обробки в середньому на 40% без втрати точності. Отримані результати свідчать про покращення коефіцієнта силуету до 0,73 та підвищення стабільності кластерного поділу порівняно з базовими алгоритмами.

Практична цінність роботи полягає у можливості інтеграції розробленої моделі у системи аналітики великих даних для транспортної, економічної та екологічної сфер. Подальші дослідження доцільно спрямувати на вдосконалення механізмів самоорганізації кластерів, розширення адаптивності моделі до потокових даних і поєднання з технологіями глибинного навчання.

Список використаних джерел

1. Бойко Н. І., Ткачик О. А. Алгоритми та методи кластеризації для різноманітних даних. Науковий вісник Ужгородського університету. Серія «Математика і інформатика». 2023;42(1):129–147. https://doi.org/10.24144/2616-7700.2023.42(1).129-147.
2. Ammar Alramahee, Fahad Ghalib A Survey of Clustering Algorithms for Determining Optimal Locations of Distributed Centers. Basrah Researches Sciences. 2024;50(2):318–332. https://doi.org/10.56714/bjrs.50.2.26.

3. Oyewole G. J., Thopil G. A. Data clustering: application and trends. Artificial Intelligence Review. 2023;56:6439–6475. https://doi.org/10.1007/s10462-022-10325-y.
4. Барченко Н., Любчак В., Великодний Д. Вибір метода кластеризації з метою аналізу показників цифрових трансформацій регіонів України. Інформаційні технології та суспільство. 2023;2(8):6–17. https://doi.org/10.32689/maup.it.2023.2.1.
5. Усатенко М. В. Методи виявлення аномалій у масивах багатовимірних даних. Збірник наукових праць Харківського національного університету радіоелектроніки. 2024. URL: https://openarchive.nure.ua/bitstreams/15148cad-14b3-47db-9f0f-e7d705fcf/download.
6. Alzubaidi L., Zhang J., Humaidi A. J., та ін. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data. 2021;8:53. https://doi.org/10.1186/s40537-021-00444-8.
7. Dafir Z., Lamari Y., Slaoui S. C. A survey on parallel clustering algorithms for Big Data. Artificial Intelligence Review. 2021;54:2411–2443. https://doi.org/10.1007/s10462-020-09918-2.
8. Zhou S., Xu H., Zheng Z., Chen J., Li Z., Bu J., Wang X., Zhu W., Ester M. A Survey on Deep Clustering: Taxonomy, Challenges, and Future Directions. arXiv preprint. 2022. URL: https://arxiv.org/abs/2206.07579.
9. Hu L., Jiang M., Dong J., Liu X., He Z. Interpretable Clustering: A Survey. arXiv preprint. 2024. URL: https://arxiv.org/abs/2409.00743.
10. Ткачик О. А. Методи та засоби кластеризації різнотипових даних: дис. … доктора філософії: 122 – Комп'ютерні науки. Львів: Національний університет «Львівська політехніка»; 2023. 187 с. URL: https://lpnu.ua/sites/default/files/2023/radaphd/25194/disertaciya-metodi-ta-zasobi-klasterizacii-riznotipovikh-danikh-tkachik-o-1-1.pdf
11. .Чорна О. С., Дідик П. Ю., Тітов С. В., Тітова О. В. Використання алгоритмів кластеризації для автоматизації планування маршрутів у задачах маршрутизації перевезень. Системи обробки інформації. 2024;1(176):115–123. https://doi.org/10.30748/soi.2024.176.14.
12. Юрчишена Л. В. Кластеризація університетів та їх економічна модель на основі показників фінансової стійкості. Економічна модель розвитку університетів. 2023:73–86. URL: https://science.iea.gov.ua/wp-content/uploads/2023/10/6_Yurchyshena_324_2023_73-86.pdf.
13. Батюк Т. М., Досин Д. Г. Інтелектуальна система кластеризації користувачів соціальних мереж на основі аналізу тональності даних. Інформаційні системи та мережі. 2023;13:121–140. https://doi.org/10.23939/sisn2023.13.121.
14. Kharlamova G., Chernyak O. Cluster analysis of Ukrainian regions regarding the level of investment attractiveness. ICTERI Proceedings. 2021;2:230–241. URL: https://icteri.org/icteri-2021/proceedings/volume-2/202110401.pdf.
15. Artioli P., Maci A., Magrì A. A comprehensive investigation of clustering algorithms for User and Entity Behavior Analytics. Frontiers in Big Data. 2024;7:1375818. https://doi.org/10.3389/fdata.2024.1375818.
16. Xue J., Xing L., Wang Y., та ін. A comprehensive survey of fast graph clustering. Vicinagearth. 2024;1:7. https://doi.org/10.1007/s44336-024-00008-3.
17. Saeed M. M., Al Aghbari Z., Alsharidah M. Big data clustering techniques based on Spark: a literature review. PeerJ Computer Science. 2020;6:e321. https://doi.org/10.7717/peerj-cs.321.
18. Kohavi R. Census Income [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5GP7S.

Weigang Ganna
Candidate of Engineering Sciences, Associate Professor, Associate Professor of the Department of Computer Science,
National University of Life and Environmental Sciences of Ukraine
ORCID: https://orcid.org/0000-0002-2082-2322
E-mail: weigang.ganna@nubip.edu.ua

Naurynskyi Yurii
Assistant, Department of Computer Science,
National University of Life and Environmental Sciences of Ukraine
ORCID: https://orcid.org/0009-0004-6416-8635
E-mail: yu.naurynskyi@nubip.edu.ua

Myronchuk Kateryna
Senior Lecturer, Department of Computer Systems, Networks and Cybersecurity,
National University of Life and Environmental Sciences of Ukraine
ORCID: https://orcid.org/0000-0001-6764-3746
E-mail: k.komar@nubip.edu.ua

HYBRID CLUSTERING OPTIMIZATION MODEL FOR INTELLIGENT DECISION SUPPORT SYSTEMS

Abstract. *The article presents a comprehensive approach to optimizing clustering algorithms within decision support systems (DSS) in big data environments. It analyzes issues of scalability, computational complexity, and result stability that are typical of classical methods such as K-Means, DBSCAN, and Agglomerative Clustering. An improved hybrid algorithm, K-Means++ Hybrid, is proposed, combining parallel computing mechanisms, adaptive parameter tuning, and dynamic control of the iterative search process. The methodological foundation of the research is based on systems analysis, mathematical modeling, and experimental testing using datasets from the UCI Repository and GPU acceleration technologies (CUDA). Experimental results confirm that the proposed approach reduces clustering execution time by approximately 43% compared to baseline algorithms, while increasing the silhouette coefficient to 0.73 and reducing CPU energy consumption by 20–25%. The resulting model demonstrates high robustness when processing heterogeneous datasets and can be integrated into systems for traffic flow analysis, financial risk assessment, and environmental monitoring. The developed approach provides a foundation for building adaptive intelligent data analysis modules that support scalability, result interpretability, and real-time operation in streaming analytics systems. Future research should focus on integrating hybrid clustering with deep learning models and evolutionary optimization algorithms.*

Keywords: *Clustering, Big Data, Decision Support Systems, Algorithm Optimization, Parallel Computing, Hybrid Models, Machine Learning.*