

ВИСНОВОК

**про наукову новизну, теоретичне та практичне значення результатів
дисертації Кириченка Євгена Олександровича на тему:
“ Оптимізація структури гетерогенних даних в Big Data ”,
поданої на здобуття ступеня доктора філософії
за спеціальністю 121 – Інженерія програмного забезпечення
з галузі знань 12 – Інформаційні технології**

1. Обґрунтування вибору теми дослідження та її зв'язок із планами наукових робіт Університету.

Сучасний етап інформатизації характеризується стрімким накопиченням великих масивів гетерогенних, різномірних та слабо документованих даних. Через їхню структурну складність виникає необхідність у розробці методів аналізу (класифікації, порівняння, кластеризації), які не потребують оперування повними наборами даних. Замість цього акцент зміщується на створення компактних та інтерпретованих представлень, що зберігають ключові характеристики вихідної інформації.

Попри значний прогрес у методах аналізу табличних даних, досі залишаються невирішеними проблеми автоматичної ідентифікації типів змінних у слабо структурованих наборах, побудови гетерогенних компактних репрезентацій та розробки універсальних метрик порівняння. Важливим і перспективним напрямом є створення графових моделей подібності датасетів, що забезпечить можливість використання спектрального аналізу для виявлення кластерів та прихованих залежностей.

Вибір теми дослідження зумовлений гострою потребою в ефективному аналізі гетерогенних даних. Сьогодні провідні вчені активно досліджують методи інтеграції структурованих і неструктурованих ознак за допомогою глибинного навчання (D. Zhang та ін.), а також класифікацію характеристик за змішаними наборами даних (Н. Шаховська, І. Загородний). Важливими науковими викликами залишаються оптимізація простору ознак для побудови

компактних моделей (J. Cui та ін.) та застосування ансамблевих і статистичних підходів для аналізу багатовимірних процесів (Y. Vyklyuk, S. Malinović-Milićević, A. Ahmad та ін.).

З огляду на це, розробка теоретичних засад та інформаційної технології, що охоплює класифікацію змінних, формування компактних дескрипторів (Compact Data Representation, CDR), обчислення узагальненої структурної метрики подібності (Data Information Structure Similarity, DISS), побудову графів і дерев подібності та їх реалізацію у хмарному середовищі, є своєчасною та актуальною науково-прикладною задачею.

Мета і завдання дисертаційного дослідження. Метою дослідження є розробка уніфікованої типо-орієнтованої інформаційної системи для ефективного порівняння гетерогенних табличних наборів даних з невідомих або слабо документованих джерел за допомогою створення компактних представлень даних з використанням зваженої метрики структурної подібності даних та побудова масштабованої хмарної інформаційної системи для формування графів та дерев подібності корпусів даних з мінімізацією витрат на зберігання.

Об'єкт дослідження. Процес порівняння гетерогенних табличних наборів даних зі слабо документованих джерел.

Предмет дослідження. Моделі та методи класифікації змінних, створення компактних представлень гетерогенних табличних наборів даних, побудова графів подібності для ефективного порівняння та структурування корпусів даних за мінімальних припущень щодо схеми, кластеризація корпусів даних на основі спектрального аналізу матриці подібності.

Методи дослідження. У дисертації застосовано методи теорії ймовірностей та математичної статистики для моделювання властивостей змінних та оцінювання статистичних характеристик компактних представлень; спектральні методи лінійної алгебри для дослідження структурних властивостей графа подібності; методи машинного навчання для побудови уніфікованої метрики структурної подібності інформації даних; аналітичні та

числові методи для дослідження властивостей запропонованих моделей; хмарні технології AWS для реалізації масштабованої інформаційної технології.

Дисертаційне дослідження виконано відповідно до планів науково-дослідницьких робіт:

- кафедри програмного забезпечення комп'ютерних систем Чернівецького національного університету імені Юрія Федьковича за держбюджетною тематикою: «Дослідження, моделювання та розробка програмного забезпечення складних динамічних систем» (Державний реєстраційний номер 0121U109232);
- кафедри математичних проблем управління і кібернетики Чернівецького національного університету імені Юрія Федьковича за держбюджетною тематикою: «Інформаційні технології в аспекті сучасних задач прийняття рішень» (Державний реєстраційний номер 0121U109159).

2. Формулювання наукового завдання, нове розв'язання якого отримано в дисертації.

Наукові завдання, розв'язання яких отримано у дисертації, полягають у наступному:

- дослідити сучасні методи компактного подання даних, класифікації змінних та метрик подібності;
- розробити процедуру конструювання компактних представлень даних для різних типів змінних (факторних (таблиці частот), часових (гістограми з фіксованими інтервалами), числових (вектори моментів) та рядкових (вектори TF-IDF));
- створити зважену, типоспецифічну метрику подібності, що об'єднує відстані між гетерогенними представленнями через навчені або встановлені користувачем ваги;
- розробити теоретичне підґрунтя кластеризації даних на основі спектрального аналізу матриці подібності, побудувати граф подібності (та пов'язане з ним дерево) для наборів даних, використовуючи лише компактні, а не повні копії даних;

– розробити інформаційну технологію для обробки великих корпусів табличних даних з оптимізацією витрат на зберігання із реалізацією в AWS-середовищі (S3, EMR, Glue, Athena, Airflow);

– виконати порівняльні експерименти для виявлення схожих наборів даних та оцінки структурної якості корпусу.

3. Наукові положення, розроблені особисто дисертантом, та їх новизна:

- проведено ґрунтовний аналітичний огляд сучасних підходів до кластеризації великих графових структур та ролі випадкових матриць у моделюванні складних мереж, показавши обмеженість класичних підходів у випадку гетерогенних графів із незалежними та мінливими вагами;
- отримано теоретичні результати для класу випадкових матриць, елементи яких характеризуються розподілами з важкими хвостами. Сформульовано гіпотези, які формують математичний каркас для опису кластеризації графів та гарантують коректність переходу до блочної моделі;
- на основі теоретичних результатів побудовано критерій визначення оптимальної кількості кластерів k_{opt} ;
- запропоновано новий підхід побудови CDR та математичні моделі побудови компактних дескрипторів для числових, категоріальних, текстових і часових змінних;
- запропоновано та реалізовано систему метрик структурної подібності DISS, до складу якої включено відстані Вассерштейна, Геллінгера, повної варіації, косинусну метрику та спеціальні метрики, що забезпечує високу чутливість до змін форми розподілу;
- модифіковано підхід до кластеризації табличних даних на основі зваженого графа метрик DISS, поєднавши спектральні властивості випадкових матриць з практичними алгоритмами кластеризації на компактних дескрипторах;

- розроблено повноцінну інформаційну систему для побудови CDR, обчислення DISS, аналізу дрейфу та кластеризації даних, що базується на модульних компонентах і підтримує обробку великих хмарних даних;
- розроблено хмарну модель розгортання інформаційної системи на основі Apache Spark та AWS EMR/EMR Serverless, що забезпечує автоматичне масштабування та зниження витрат на обчислення;
- проведено повний комплекс експериментальних досліджень, які демонструють, що система CDR/DISS дозволяє точно виявляти зміни в структурі даних навіть при значному зменшенні обсягів інформації.

На всіх етапах досліджень нових задач і розробки нових методів дисертант брав активну участь. Ним здійснено ґрунтовний аналіз отриманих результатів та сформовано висновки до кожного розділу дисертації.

4. Обґрунтованість і достовірність наукових положень, висновків і рекомендацій, які захищаються.

Достовірність наукових положень та висновків обґрунтовані тим, що для розробки програмного забезпечення використані сучасні методи та середовище розробки; програмне забезпечення протестоване стандартними методами тестування; отримані за допомогою розробленої інформаційної технології дані не суперечать загальноприйнятим міркуванням і принципам, а теоретичні положення, розроблені дисертантом із застосуванням добре апробованих теоретичних методів, перевірені та корелюють із класичними методами визначення оптимальної кількості кластерів у великих графових структурах.

Дисертація складається із вступу, чотирьох розділів, висновків до розділів, загальних висновків, переліку використаних джерел та трьох додатків. Дисертаційна робота є самостійною науковою працею, має теоретичне та прикладне значення. Усі висновки, рекомендації та положення, що характеризують наукову новизну дослідження, сформульовані особисто автором дисертації.

Публікації (дві статті у виданнях, що індексуються у наукометричній базі даних SCOPUS, три статті у наукових фахових виданнях України, 1 стаття, яка

додатково відображає результати дисертації) у наукових журналах за темою дисертації висвітлюють проблематику та основні положення наукової роботи. Апробація основних наукових результатів відбулася у формі доповідей на 6 Міжнародних науково-практичних конференціях.

5. Рівень теоретичної підготовки здобувача та рівень його обізнаності з результатами наукових досліджень інших науковців.

Здобувач має високий рівень теоретичної та практичної підготовки в галузі інформаційних технологій, що дозволяє йому ефективно вирішувати комплексні наукові завдання. Здобувач продемонстрував високий рівень обізнаності з науковими досягненнями інших учених стосовно досліджень дисертаційної роботи. Проведений аналіз наукових досягнень у цій сфері дозволив здобувачу значно розширити розуміння сучасних тенденцій та вдосконалити існуючі підходи для досягнення нових результатів.

Особистий внесок здобувача у розв'язанні конкретних наукових завдань також досить високий. Зокрема:

- дисертант брав участь в обговоренні постановки задачі, визначенні мети роботи, виборі методів досліджень та у підготовці матеріалів до публікації у наукових журналах, представленні на наукових конференціях із доповіддю отриманих результатів;
- займався розробкою, тестуванням та вдосконаленням програмного забезпечення;
- виконував експериментальні дослідження, основні розрахунки та аналіз отриманих результатів;
- займався питаннями застосування теоретичного результату до визначення оптимальної кількості кластерів моделей Big Data, включно з графами, що мають розподіли з важкими хвостами;
- проводив теоретичне дослідження та здійснював доведення основних теоретичних тверджень дисертаційного дослідження.

6. Наукове та практичне значення роботи.

Важливе наукове значення дисертаційної роботи полягає в тому, що результати теоретичних досліджень, а саме розвитку теорії графових досліджень, сформульовані та доведені теореми, можуть використовуватися для подальших досліджень у цій галузі, адже на сьогодні немає універсального способу проведення процесу кластеризації у Big Data та гетерогенних мережевих системах. А застосування методу компактних представлень даних на основі типо-специфічних зведень дозволяє зменшити обсяги даних для зберігання та передачі при збереженні достатньої інформації для порівняння наборів даних.

Практичне значення дисертаційної роботи полягає в тому, що розроблена автором інформаційна технологія для порівняння гетерогенних табличних наборів даних, процедура конструювання компактних представлень даних, метрики обчислення подібності та методи побудови структур корпусів даних можуть використовуватися для подальшого вивчення інформаційних систем великого масштабу.

7. Використання результатів роботи.

Запропоновані дисертантом підходи до CDR та метрики DISS для порівняння табличних даних, розроблений метод кластеризації на основі структурних відстаней та запропоновані методи інтеграції кластерного аналізу у аналітичні модулі компанії використовуються у роботі компаній ТОВ «Кодерс ПРО» та ТОВ «Палетний сервіс». А результати теоретичних та практичних досліджень використовуються у навчальному процесі кафедри математичних проблем управління і кібернетики та кафедри програмного забезпечення комп'ютерних систем Чернівецького національного університету імені Юрія Федьковича.

8. Повнота викладу матеріалів дисертації в публікаціях та особистий внесок здобувача в публікації, виконані у співавторстві.

Особистий внесок здобувача в публікації такий, який вказаний у пункті 5 цього висновку.

Результати перевірки тексту дисертації з використанням антиплагіатної системи TURNITIN показав 3% схожості з джерелами з Інтернету. Робота відповідає принципам академічної доброчесності.

Основні положення і висновки дисертаційної роботи викладені у 6 наукових працях. Зокрема, три з них індексована у наукометричній базі Scopus, 3 наукові праці – у виданнях, що включені до переліку наукових фахових видань України (категорія Б). Кількість наукових праць, які додатково відображають наукові результати дисертації – 1. Матеріали дисертації доповідалися і опубліковані у 6 тезах доповідей на Міжнародних та Всеукраїнських наукових конференціях.

Результати дисертації повною мірою викладені в зазначених публікаціях.

**Наукові праці, в яких опубліковані
основні наукові публікації дисертації:**

***Наукові праці у виданнях, включених до переліку
наукових фахових видань України та проіндексованих у наукометричній
базі даних Scopus:***

1. Кириченко О., **Кириченко Є.** Асимптотичні властивості випадкових матриць = Asymptotic Properties of Randommatrices. *Вісник Київського національного університету імені Тараса Шевченка. Фізико-математичні науки.* 2024. 78(1). С. 41–44. (**Scopus**) <https://doi.org/10.17721/1812-5409.2024/1.7> (*Кириченко О. – концептуалізація, формальний аналіз; Кириченко Є. – аналіз джерел, підготовка огляду літератури*).

***Наукові праці у періодичних наукових виданнях,
проіндексованих у наукометричній базі даних Scopus:***

2. Malyk I. V., **Kyrychenko Y.**, Gorbatenko M., Lukashiv T. Data Optimization through Compression Methods Using Information Technology. *International Journal of Information Technology and Computer Science.* 2025. Vol.17, No.5. P. 84-99. (**Scopus, Q3**) <https://doi.org/10.5815/ijitcs.2025.05.07> (*Малик І. – концептуалізація, методологія, формальний аналіз, написання та підготовка оригінальної чернетки, написання, рецензування та*

редагування, візуалізація, авторський нагляд; Кириченко Є. – концептуалізація, методологія, формальний аналіз, дослідження, написання та підготовка оригінальної чернетки, написання-рецензування та редагування, візуалізація; Горбатенко М. – методологія, формальний аналіз; Лукашів Т. – концептуалізація, написання, рецензування та редагування, авторський нагляд, адміністрування проєкту).

**Наукові праці у виданнях, включених до переліку
наукових фахових видань України:**

3. **Кириченко Є. О.**, Кириченко О. Л. Побудова моделей нейронних мереж у хмарному середовищі засобами AWS EMR. *Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія : Технічні науки.* 2024. 35(74), №6. С. 75-79. <https://doi.org/10.32782/2663-5941/2024.6.2/11> (Кириченко Є. – теоретична та практична розробка положень, програмна реалізація, редагування рукопису; Кириченко О. – постановка задачі, керівництво дослідженням, рецензування та редагування).
4. **Kyrychenko Y.**, Malyk I. Information Technology for Data Compression and Transformation by Means of Apache Airflow and Amazon EMR. *Security of Infocommunication Systems and Internet of Things.* 2025. Vol. 3, no. 1. P. 01004. <https://doi.org/10.31861/sisiot2025.1.01004> (Кириченко Є. – проєктування архітектури системи, програмне забезпечення, розробка та тестування програмного забезпечення, написання та редагування; Малик І. – загальне керівництво, концептуалізація).
5. **Кириченко Є.**, Малик, І. Один результат кластеризації на основі випадкових матриць. *Information Technology: Computer Science, Software Engineering and Cyber Security.* 2025. №3. С. 257-270. <https://doi.org/10.32782/IT/2025-3-29> (Кириченко Є. – проведення дослідження, редагування рукопису; Малик І. – постановка задачі, визначення загальної схеми досліджень, обговорення результату).

**Наукові праці, які засвідчують
апробацію матеріалів дисертації:**

6. **Кириченко Є. О.**, Малик І. В. Використання подвійних стохастичних моделей в BIG DATA. *Проблеми інформатики та комп'ютерної техніки (ПІКТ – 2023)* : праці XII Міжнародної науково-практичної конференції, м. Чернівці, 10 – 12 листопада 2023 р. Чернівці : Черн. нац. ун-т, 2023. С. 33–35. (Кириченко Є. – огляд літературних джерел, написання, рецензування та редагування; Малик І. – постановка задачі, рецензування).
7. **Кириченко Є. О.**, Малик І. В. Моделі і процеси при дослідженні подвійних стохастичних систем. *Інформаційні технології: наука, техніка, технологія, освіта, здоров'я* : тези доп. XXXII Міжнар. наук.-практ. конф. MicroCAD-2024, 22–25 травня 2024 р. / за ред. проф. Сокола Є.І. Харків : НТУ «ХПІ». С. 1294. (Кириченко Є. – проведення дослідження, написання рукопису; Малик І. – концептуалізація, методологія, загальне керівництво).
8. **Kyrychenko Y.**, Malyk I. Leveraging AWS EMR for Scalable and Efficient Neural Network Deployment in Cloud Computing. *The 13 th International Conference on Electronics, Communications and Computing's (IC ECCO)* : Materials of the Intern. Conf., (Chisinau, Moldova, 17–18 October, 2024.). Chisinau : Technical University of Moldova, 2024. P. 128-129. (Кириченко Є. – проведення досліджень, аналіз та опис результатів; Малик І. – постановка задачі, обговорення результатів).
9. **Кириченко Є. О.**, Малик І. В. Порівняльний аналіз економічної ефективності AWS ELASTIC MAPREDUCE та EC2 для розгортання нейронних мереж. *Проблеми інформатики та комп'ютерної техніки (ПІКТ–2024)* : праці XIII Міжнар. наук.-практ. конф., м. Чернівці, 01–03 листопада 2024 р. Чернівці : Черн. нац. ун-т, 2024. С. 40–42. (Кириченко Є. – збір та аналіз даних, обговорення результатів; Малик І. –

концептуалізація дослідження, розробка методології, аналіз результатів).

10. **Кириченко Є.**, Малик І. Інформаційна технологія структурного аналізу табличних наборів даних для виявлення аномалій та невідповідностей. *Science in the Context of Modern Challenges and Prospects* : Materials of the Intern. Conf., 21 травня 2025 р. (м. Утрехт, Нідерланди). 2025. С. 196-197. (Кириченко Є. – проведення дослідження, аналіз та опис результату; Малик І. – визначення загальної схеми дослідження).
11. Малик І. В., **Кириченко Є. О.** Підхід до порівняння гетерогенних табличних даних на основі методів компресії. *Проблеми інформатики та комп'ютерної техніки* : праці XIV Міжнародної науково-практичної конференції (ПІКТ – 2025), присвяченої 85-й річниці від дня народження проф. Кириченка М.Ф. м. Чернівці, 13–15 лист. 2025. Чернівці : Технодрук, 2025. С. 124-126. (Кириченко Є. – концептуалізація, методологія, формальний аналіз, дослідження, написання; Малик І. – постановка задачі, загальне керівництво).

**Наукові праці, які додатково відображають
наукові результати дисертації:**

12. **Kyrychenko Y. O.**, Kyrychenko O. L., Filipchuk O. I. The Study Of Complex Networks Using Visualization Tools. *Science and technology today*. «Technics» series. 2024. № 10(38). С. 81-93. [https://doi.org/10.52058/2786-6025-2024-10\(38\)-81-92](https://doi.org/10.52058/2786-6025-2024-10(38)-81-92) (Кириченко Є. – концептуалізація, методологія, формальний аналіз, дослідження, підготовка ілюстративних матеріалів; Кириченко О. – концептуалізація, обговорення результатів; Філіпчук О. – обговорення результатів, рецензування та редагування).

9. Апробація матеріалів дисертації

Апробація матеріалів дисертації здійснювалася на таких конференціях та семінарах:

- «Проблеми інформатики та комп'ютерної техніки (ПІКТ)» (Чернівці, 2023-2025),

- «Інформаційні технології: наука, техніка, технологія, освіта, здоров'я (MicroCAD)» (Харків, 2024),
- The 13th International Conference on Electronics, Communications and Computing's (IC ECCO) (м. Кишинів, Молдова, 2024),
- «Science in the Context of Modern Challenges and Prospects» (м. Утрехт, Нідерланди, 2025),
- наукові семінари кафедр математичних проблем управління і кібернетики та програмного забезпечення комп'ютерних систем Чернівецького національного університету імені Юрія Федьковича.

10. Оцінка мови і стилю дисертації.

Дисертація написана чіткою мовою, відповідає критеріям науковості, забезпечуючи логічність, послідовність і об'єктивність викладення результатів дослідження. Зазначене свідчить про відповідність вимогам, що висуваються до праць такого рівня.

11. Відповідність змісту дисертації спеціальності з відповідної галузі знань, з якої вона подається до захисту.

Зміст дисертації відповідає чинним вимогам до оформлення дисертації, встановленим освітньо-науковою програмою «Інженерія програмного забезпечення» галузі знань 12 «Інформаційні технології», спеціальності 121 – Інженерія програмного забезпечення.

12. Дотримання нормативних вимог щодо оформлення дисертації.

Нормативні вимоги щодо оформлення дисертації дотримані повністю.

13. Рекомендації дисертації до захисту.

Дисертаційна робота Кириченка Євгена Олександровича «Оптимізація структури гетерогенних даних в Big Data», подана на здобуття ступеня доктора філософії (PhD) у галузі знань 12 – Інформаційні технології за спеціальністю 121 – Інженерія програмного забезпечення, за її актуальністю, науково-технічним рівнем, новизною постановки та розв'язання проблем, практичним значенням отриманих результатів відповідає вимогам пунктів 6, 7, 8, 9

«Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого Постановою Кабінету Міністрів України від 12 січня 2022 р. №44 (зі змінами, внесеними згідно з Постановою Кабінету Міністрів України №507 від 03.05.2024 р.).

За результатами публічної презентації результатів дисертації та їх обговорення на засіданні кафедри програмного забезпечення комп'ютерних систем Навчально-наукового інституту фізико-технічних та комп'ютерних наук Чернівецького національного університету імені Юрія Федьковича 03 березня 2026 року дисертацію Кириченка Євгена Олександровича рекомендовано до захисту в разовій спеціалізованій вченій раді для здобуття ступеня доктора філософії (PhD) з галузі знань 12 – Інформаційні технології за спеціальністю 121 – Інженерія програмного забезпечення.

Голова засідання

Катерина ГАЗДЮК

доктор філософії, доцент,
завідувач кафедри програмного
забезпечення комп'ютерних систем
Чернівецького національного університету
імені Юрія Федьковича

09.03.26.

