УДК: 004.855.5:519.237.8 https://doi.org/10.15407/jai2026.01.071

О. В. Іващенко[1], **С. С. Федін**[2]
[1,2]Національний транспортний університет, Україна
вул. Омеляновича-Павленка, 1, м. Київ, 01010
[1]alexander.ivashchenko@gmail.com
[2]sergey.fedin1975@gmail.com
[1]https://orcid.org/0009-0007-5470-9137
[2]https://orcid.org/0000-0001-9732-632X

ЗАБЕЗПЕЧЕННЯ ТОЧНОСТІ ТА СТАБІЛЬНОСТІ КЛАСТЕРИЗАЦІЇ БАГАТОВИМІРНИХ ДАНИХ МЕТОДОМ САМООРГАНІЗАЦІЙНИХ КАРТ КОХОНЕНА НА ОСНОВІ АВТОМАТИЧНОЇ РЕДУКЦІЇ ДАНИХ

Анотація. У статті досліджено метод підвищення точності та стабільності кластеризації багатовимірних даних шляхом поєднання самоорганізаційних карт Кохонена з процедурою автоматичної редукції ознак. Запропоновано метод вилучення малозначущих ознак з використанням алгоритму міжквартильного розмаху, інтегрованого у створений програмний додаток для побудови SOM. Проведено порівняння результатів кластеризації до та після редукції за незмінних параметрів навчання і фіксованого значення seed – параметра, що визначає початковий стан генератора випадкових чисел та забезпечує відтворюваність ініціалізації вагових коефіцієнтів нейронів. Засвідчено, що вилучення ознак із низькою варіативністю не порушує топологічну структуру SOM, зберігає стабільність кластерного поділу та зменшує обчислювальні витрати. Підтверджено високий ступінь відповідності між даними до та після редукції на основі статистичного аналізу. Встановлено, що запропонований метод підвищує ефективність SOM і є придатним для аналізу великих багатовимірних вибірок.

Ключові слова: SOM, редукція даних, малозначущі ознаки, кластеризація, міжквартильний розмах, програмний додаток.

O. Ivashchenko[1], **S. Fedin**[2]
[1,2]National transport university, Ukraine
st. Omelianovycha-Pavlenka, 1, Kyiv, 01010
[1]alexander.ivashchenkoo@gmail.com
[2]sergey.fedin1975@gmail.com
[1]https://orcid.org/0009-0007-5470-9137
[2]https://orcid.org/0000-0001-9732-632X

ENSURING ACCURACY AND STABILITY OF MULTIDIMENSIONAL DATA CLUSTERING USING KOHONEN SELF-ORGANIZING MAPS BASED ON AUTOMATIC DATA REDUCTION

Abstract. The article investigates a method for improving the accuracy and stability of multidimensional data clustering by combining Kohonen self-organizing maps with an automatic feature reduction procedure. A method for removing insignificant features using an interquartile range algorithm integrated into a software application for building SOMs is proposed. A comparison of the clustering results before and after reduction is performed with unchanged training parameters and a fixed seed value, a parameter that determines the initial state of the random number generator and ensures the reproducibility of the initialization of the neuron weight coefficients. It has been proven that the removal of features with low variability does not violate the topological structure of SOM, preserves the stability of cluster separation, and reduces computational costs. A high degree of correspondence between the data before and after reduction has been confirmed based on statistical analysis. It has been established that the proposed method increases the efficiency of SOM and is suitable for analyzing large multidimensional samples.

Keywords: SOM, data reduction, insignificant features, clusterization, interquartile range, software application.

Вступ

У сучасних умовах стрімкого зростання обсягів даних постає необхідність ефективних методів їх аналізу, візуалізації та узагальнення. Одним із ефективних інструментів для дослідження багатовимірних даних є самоорганізаційні карти Кохонена (Self-Organizing Maps, SOM). SOM дозволяє здійснювати проєкцію багатовимірних вхідних даних у двовимірний простір, зберігаючи при цьому топологічні

властивості первинної вибірки. Такий метод дає змогу не лише візуалізувати структуру даних, а й виявляти приховані закономірності, кластери та взаємозв'язки між ознаками, що є особливо важливим для аналізу складних систем у таких галузях, як телекомунікаційні системи, CRM-системи, фінансове моделювання, біомедичні дослідження тощо.

При цьому ефективність побудови самоорганізаційних карт суттєво залежить від якості та релевантності вхідних ознак. У випадку, коли вихідний простір ознак містить надлишкові або малоінформативні дані, модель може зазнавати перенавчання, що спотворює результати кластеризації та візуалізації. Крім того, надмірна кількість вимірів ускладнює процес навчання карти та підвищує обчислювальні витрати. Тому постає завдання не лише у зменшенні розмірності даних, яке природно здійснюється SOM під час побудови двовимірної проєкції, але й у скороченні кількості вхідних ознак до початку навчання SOM моделі.

У зв'язку з цим актуальною науковою задачею є потреба вдосконалення методів побудови самоорганізаційних карт Кохонена, які враховують попередню редукцію вхідних ознак як важливу частину аналізу даних. При цьому одним із ключових процесів є створення програмного інструмента, який дозволяє поєднати процедуру відбору найбільш інформативних ознак із процедурою побудови карти Кохонена. Це дає змогу отримати більш точну візуалізацію структур даних, спростити їх обробку та підвищити інтерпретованість результатів для подальшого використання на практиці. Такий метод дозволить краще адаптувати SOM до особливостей конкретної предметної області та підвищити достовірність отриманих результатів.

Огляд літератури

Принцип роботи SOM ґрунтується на тому, що кожен вхідний об'єкт, описаний у просторі ознак, проєктується на двовимірну решітку нейронів, де кожен нейрон має власний вектор вагових коефіцієнтів, який відображає характерні властивості певної групи схожих об'єктів. SOM виконує функцію зменшення розмірності даних, перетворюючи великі багатовимірні набори спостережень у компактне топологічно - впорядковане представлення. Наприклад, набір даних, що містить 5000 зразків із десятьма ознаками, після навчання відображається на карту розміром 30×30×10, тобто на десять двовимірних карт – по одній для кожної ознаки. У такому разі 5000 вихідних точок замінюються 900 нейронами, що суттєво скорочує обсяг інформації, але зберігає просторові взаємозв'язки між подібними об'єктами [1].

У дослідженнях С. Забєліна [2] та М. Вайсьєра [3] SOM розглядається як один із ключових методів нелінійного зменшення розмірності (NLDR, Nonlinear Dimensionality Reduction). Автори систематизують методи до виявлення низьковимірних структур у складних даних, порівнюючи SOM із MDS, Isomap, GTM та іншими алгоритмами NLDR. Визначено, що SOM поєднує здатність зменшувати розмірність із можливістю зберігати локальні й глобальні відносини між точками, що робить цей метод особливо ефективним для дослідницького аналізу багатовимірних вибірок.

У статті К. Агбоки [4] SOM використано для попереднього групування екологічних і біологічних показників, що дозволяє виокремити інформативні патерни та зменшити вплив малозначущих ознак, після чого скорочений набір даних подається на згорткову нейронну мережу (CNN, Convolutional Neural Network). Така двоетапна схема SOM+CNN забезпечує більшу точність класифікації порівняно з використанням лише CNN, оскільки SOM ефективно виконує редукцію та структуризацію даних. Подібний метод реалізовано у роботі С. Салап-Айджа [5], де SOM застосовано для зменшення розмірності просторових показників і спрощення аналізу великої кількості карт впливу. Визначено, що надлишок вхідних ознак ускладнює інтерпретацію результатів та призводить до перенасичення моделі, тоді як SOM дозволяє згрупувати схожі

просторові патерни у компактній двовимірній формі.

Ефективність SOM у відборі релевантних ознак підтверджують також результати дослідження Р. Г. Сілви та С. Дж. Вілкокса [6], які поєднали SOM із просторовою статистикою для оцінки сенсорних ознак у задачах моніторингу зносу інструмента. Поєднання SOM зі статистичними методами дозволяє визначити внесок кожної ознаки навіть за наявності шуму, що демонструє придатність використання SOM у технічних системах із великою кількістю сенсорних даних.

У роботі Л. Хасефа [7] показано, що точність SOM знижується зі збільшенням кількості ознак, тому перед навчанням доцільно застосовувати спеціалізовані нейронні мережі для їх вилучення. Порівняння методів на основі розріджених згорткових автокодерів і спайкових нейронних мереж показало підвищення точності кластеризації SOM більш ніж на 6 %, що підтвердило ефективність інтеграції методів автоматичного вилучення ознак із SOM для аналізу складних даних.

У роботі С. Фаня [8] запропоновано вдосконалення SOM шляхом використання ініціалізації ваг методом головних компонентів (PCA, Principal Component Analysis) і механізму «віртуального переможного нейрона». Такий метод покращує топологічну узгодженість карти та підвищує точність кластеризації на різних наборах даних. Подібне дослідження провів Е. Старкі [9], у якому представлено метод Feature Weighted Growing Self-Organising Map (FWGSOM) для локального відбору релевантних ознак на рівні окремих класів. Такий метод забезпечує точніше визначення значущих ознак у межах вузлів карти та підвищує стійкість кластеризації в умовах нерівномірної інформативності ознак.

У статті К. Бенабдеслема та М. Леббаха [10] представлено метод Hl-SOM, спрямований на автоматичний відбір інформативних ознак під час навчання карти. Автори зазначають, що надлишкові або малозначущі ознаки спотворюють топологію SOM, тому процедуру редукції інтегровано безпосередньо в процес навчання. Подібний метод реалізовано в роботі О. Джейлана та Г. Ташкин Каї [11], де SOM поєднано з еволюційною оптимізацією для класифікації гіперспектральних зображень. Автори зазначають, що великі набори спектральних ознак часто є корельованими та шумовими, що ускладнює кластеризацію та посилює ефект «прокляття розмірності». Застосований еволюційний механізм забезпечує оцінювання інформативності ознак і формування мінімального релевантного набору, що підвищує точність моделі SOM.

У дослідженні Д. Пратіві [12] SOM поєднано з методами відбору ознак для класифікації зображень. Автор використовує векторні ознаки на основі гістограм яскравості та демонструє, що застосування PCA чи LSA для редукції ознак формує компактні представлення й підвищує якість кластеризації SOM. Подібний метод використано М. Хелілем [13], який інтегрував SOM у систему моніторингу якості води: аналіз SOM дав змогу виокремити найбільш інформативні фізико-хімічні ознаки та зменшити розмірність без втрати структури даних, що покращило точність і скоротило час подальшої кластеризації глибинними моделями.

У роботі З. Зана [14] запропоновано модель UDRN (Unified Dimensionality Reduction Network) – уніфіковану нейронну архітектуру для редукції ознак, яка поєднує їхній відбір із побудовою низьковимірного представлення. Модель містить дві взаємопов’язані мережі: одна відсіює нерелевантні ознаки, інша формує компактне представлення із збереженням локальної структури даних. Це забезпечує точнішу кластеризацію та кращу візуалізацію порівняно з окремими методами відбору нерелевантних ознак.

У роботах Б. Мвангі, С. Чена, Р. Россі [15-17] підкреслюється ключова роль редукції ознак у підвищенні точності, стабільності та обчислювальної ефективності моделей машинного навчання. У цих дослідженнях редукція розглядається як спосіб усунення

надлишкових, корельованих і нерелевантних ознак, що зменшують якість моделювання та загострюють проблему високої розмірності даних. Оглядові роботи описують переваги класичних методів відбору ознак – статистичних фільтрів, алгоритмів пошуку оптимальних підмножин ознак та методів, інтегрованих безпосередньо в модель. Ці методи дають змогу спростити моделі, зменшити ризик перенавчання та прискорити обробку даних.

У роботах А. Пассім'єра, А. Треліна, С. Уґіароглу [18-20] редукція даних розглядається як важливий етап обробки високовимірних вибірок, а у роботі В. Цзя, М. Сунь, Ц. Ляна та С. Хоу [21] подано огляд методів зменшення розмірності, де редукція ознак розглядається як ключовий засіб подолання «прокляття розмірності» та зниження обчислювальної складності.

Узагальнюючи проведений огляд, можна зробити висновок, що частина досліджень спрямована на використання SOM як засобу зменшення розмірності та виявлення структур у багатовимірних даних, насамперед для підвищення якості кластеризації чи візуалізації результатів [2–6, 12–14]. Інша група робіт спрямована на підвищенні точності та стабільності SOM шляхом удосконалення методів редукції ознак, зокрема через застосування автокодерів, спайкових нейронних мереж, PCA та локальних вагових моделей [7–11]. Оглядові та теоретичні праці підкреслюють важливість редукції ознак як ключової умови ефективної обробки даних у високих розмірностях і зменшення обчислювальних витрат [15–21]. Попри різноманітність методів, існуючі дослідження майже не розглядають можливість реалізації редукції ознак, інтегрованої безпосередньо у процес навчання SOM, що створює потребу в її окремому опрацюванні. Це, у свою чергу, визначає актуальність розробки методики, спрямованої на кількісне та якісне порівняння впливу редукції ознак на якість навчання самоорганізаційних карт.

Мета і завдання дослідження

Метою дослідження є забезпечення точності та стабільності кластеризації багатовимірних даних з використанням самоорганізаційних карт Кохонена та процедури автоматичної редукції даних, що забезпечує усунення малозначущих ознак та збереження топологічної структури побудованої карти.

Для досягнення поставленої мети необхідно:

– розробити метод редукції даних, спрямований на усунення малозначущих ознак вибірки;

– провести кластеризацію даних з використанням самоорганізаційних карт Кохонена до та після редукції ознак та оцінити вплив редукції на точність і якість кластеризації.

Змістовна постановка задачі

У процесі кластеризації багатовимірних даних методом SOM виникає проблема зниження точності та стабільності результатів через наявність у вхідному наборі ознак надлишкових, корельованих або малозначущих ознак. Такі ознаки не формують суттєвого внеску у побудову топологічної структури карти, проте створюють шумовий вплив і спотворюють межі кластерів, що ускладнює інтерпретацію отриманих результатів. Оскільки SOM не має вбудованого механізму автоматичного визначення інформативності ознак, постає потреба у розробці методу, який би виконував попередню редукцію даних з урахуванням їх реальної варіативності у структурі карти.

Традиційні методи відбору ознак, що ґрунтуються на фіксованих порогах чи зовнішньому налаштуванні, не завжди забезпечують узагальненість і можуть призводити до втрати важливої інформації. Тому актуальним є створення методу, здатного автоматично визначати малозначущі ознаки на основі їх статистичної варіації у структурі SOM та виконувати редукцію без порушення топології карти.

Для реалізації цього методу запропоновано інтеграцію механізму автоматичної редукції даних у процес кластеризації SOM. В основу методу покладено використання алгоритму

міжквартильного розмаху (Interquartile Range, IQR) як статистичної міри варіативності, що дозволяє виявляти ознаки з низькою дисперсією вагових коефіцієнтів у вузлах карти та виключати їх з подальшого аналізу. Таким чином, редукція виконується динамічно, без необхідності попереднього визначення порогових значень.

На основі цього методу розроблено поетапну схему (рис. 1), що включає такі основні етапи:

1. Підготовка вхідних даних з повним набором ознак.
2. Первинне навчання SOM.
3. Аналіз карти SOM із визначенням варіативності ознак алгоритмом міжквартильного розмаху.
4. Редукція даних шляхом автоматичного виключення малозначущих ознак.
5. Повторне навчання SOM на скороченому наборі ознак.
6. Порівняльний аналіз результатів кластеризації до та після редукції.

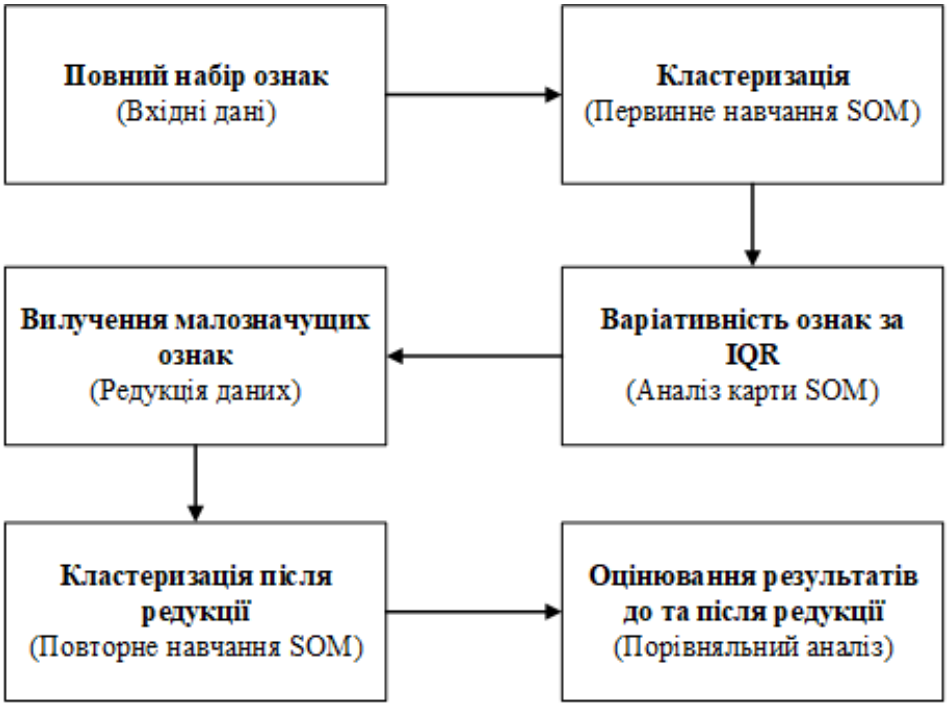


Рис. 1. Структурна схема процесу забезпечення точності та стабільності кластеризації SOM на основі процедури автоматичної редукції даних

Реалізація наведених етапів передбачає застосування спеціалізованого програмного забезпечення (СПЗ) [22, 23], що здатне здійснювати побудову самоорганізаційних карт Кохонена, обчислювати середні значення ознак для кожного сформованого кластера та містить засоби для проведення топологічного аналізу їх структури. Водночас базова функціональність СПЗ не включає механізмів автоматичної редукції ознак, застосування яких є необхідним для підвищення точності та стабільності кластеризації SOM.

З метою підвищення ефективності аналітичного процесу та розширення практичних можливостей розробленого програмного рішення запропоновано його удосконалення шляхом впровадження методу редукції даних, що забезпечує автоматичне визначення та виключення малозначущих ознак, які істотно не впливають на формування кластерів.

Матеріали та методи дослідження

Для розробки методу редукції даних у програмному додатку використано алгоритм IQR [24], який широко застосовується у статистичному аналізі для виявлення аномальних або малозмінних значень у вибірках. Такий метод дозволяє визначати ознаки, що мають низьку

варіативність, і виключати їх із подальшого аналізу без необхідності задання фіксованого порогу відсіву, оскільки поріг формується автоматично на основі статистичних характеристик вибірки. Основна ідея алгоритму полягає в тому, що ознаки з незначною дисперсією вагових коефіцієнтів у вузлах карти Кохонена мають менший вплив на процес формування кластерів і можуть бути віднесені до малозначущих.

Для кожної ознаки обчислюються статистичні характеристики, що відображають ступінь її варіативності в межах навченої карти Кохонена. Функція CalculateFeatureVariance (рис. 2) виконує збір значень вагових коефіцієнтів кожної ознаки з усіх вузлів карти, після чого визначає їх середнє значення та дисперсію. Отримана дисперсія показує, наскільки суттєво змінюється певна ознака між різними ділянками карти, тобто наскільки вона впливає на формування меж між кластерами.

```
1 reference | Oleksandr, 12 days ago | 1 author, 2 changes
public Dictionary<string, double> CalculateFeatureVariance()
{
    var featureNames = Grid[0, 0].Weights.Select(w => w.Key).ToList();
    var result = new Dictionary<string, double>();

    foreach (var feature in featureNames)
    {
        var values = new List<double>();
        for (int x = 0; x < Width; x++)
        {
            for (int y = 0; y < Height; y++)
            {
                values.Add(Grid[x, y].Weights.First(w => w.Key == feature).GetDoubleValue());
            }
        }

        double mean = values.Average();
        double variance = values.Select(v => (v - mean) * (v - mean)).Average();
        result[feature] = variance;
    }
    return result;
}
```

Рис. 2. Лістинг коду функції CalculateFeatureVariance для обчислення варіативності ознак у карті Кохонена

На основі отриманих значень дисперсій застосовується алгоритм IQR, у межах якого для сукупності дисперсій обчислюються перший квартиль (*Q1*), третій квартиль (*Q3*) та міжквартильний розмах $IQR = Q3 - Q1$. На основі цих показників визначається порогове значення $threshold = Q1 - 1.5 \times IQR$, яке задає нижню межу варіативності. Усі ознаки, дисперсія яких є нижчою за цей поріг, вважаються малозначущими. Якщо таких ознак не виявлено, додатково враховуються ті, дисперсія яких не перевищує першого квартиля (*Q1*).

Описаний метод редукції реалізовано у функції AnalyzeLowInfluenceFeatures (рис. 3), яка працює спільно з функцією CalculateFeatureVariance. На відміну від методів редукції, що передбачають видалення фіксованого відсотка ознак, запропонований механізм є динамічним і автоматично підлаштовується під структуру конкретного набору даних. Завдяки цьому редукція виконується з урахуванням реального рівня варіативності ознак, що дозволяє зберегти інформативні ознаки.

```
1 reference | Oleksandr, 12 days ago | 1 author, 2 changes
private List<string> AnalyzeLowInfluenceFeatures()
{
    if (_map == null) return new List<string>();
    var variances = _map.CalculateFeatureVariance();
    if (variances.Count == 0) return new List<string>();
    var values = variances.Values.OrderBy(v => v).ToList();
    int n = values.Count;
    double Q1 = n > 0 ? values[(int)(0.25 * (n - 1))] : 0.0;
    double Q3 = n > 0 ? values[(int)(0.75 * (n - 1))] : 0.0;
    double IQR = Q3 - Q1;
    double threshold = Q1 - 1.5 * IQR;
    // Only features with variance below threshold are considered low influence
    var lowImpact = variances.Where(kv => kv.Value < threshold).Select(kv => kv.Key).ToList();
    // If none found, fallback to Q1 (below 25th percentile)
    if (lowImpact.Count == 0)
    {
        lowImpact = variances.Where(kv => kv.Value <= Q1).Select(kv => kv.Key).ToList();
    }
    return lowImpact;
}
```

Рис. 3. Лістинг коду функції AnalyzeLowInfluenceFeatures для виявлення малозначущих ознак алгоритмом IQR

У результаті застосування запропонованого методу редукції даних, заснованого на аналізі варіативності ознак за алгоритмом IQR, формується оновлений набір даних, очищений від малозначущих ознак. Новий набір ознак використовується для повторного навчання самоорганізаційної карти Кохонена (рис. 1), що дає змогу оцінити, як редукція впливає на структуру кластерів, стабільність моделі та візуальну інтерпретованість результатів.

Для оцінки ефективності запропонованого методу редукції даних проведено експеримент із використанням SOM до та після редукції ознак. На першому етапі підготовлено вхідні дані, що містять відомості про абонентів, сформовані на основі інформації з білінгових систем телекомунікаційних компаній. Оскільки у вибірці відсутні персональні дані, що дозволяють ідентифікувати конкретних осіб, така інформація вважається анонімізованою. Вибірка складається з 1000 спостережень, що характеризуються набором ознак (табл. 1).

Таблиця 1. Перелік вхідних ознак та їх опис

№	Назва ознаки	Опис	Розмірність
1	Age	Вік абонента	роки
2	AverageMonthCosts	Середньомісячна витрата	грн
3	AverageConversationDuration	Середня тривалість розмови	хв
4	DayCallsPerMonth	Дзвінків вдень за місяць	шт./міс.
5	EveningCallsPerMonth	Дзвінків увечері за місяць	шт./міс.
6	NightCallsPerMonth	Дзвінків уночі за місяць	шт./міс.
7	OtherCitiesCalls	Дзвінки до інших міст	шт./міс.
8	OtherCountriesCalls	Дзвінки в інші країни	шт./міс.
9	ShareOfLandlineCalls	Частка дзвінків на стаціонарні телефони	%
10	NumberOfSMS	Кількість SMS	шт./міс.

Сукупність цих ознак із різною розмірністю відображає модель поведінки користувачів і є основою для подальшої кластеризації за допомогою самоорганізаційних карт Кохонена [25].

На другому етапі виконується первинне навчання самоорганізаційної карти Кохонена з попередньо підготовленим набором даних. Для ініціалізації процесу редукції ознак перед початком навчання активовано опцію Show Data Reduction (рис. 4).

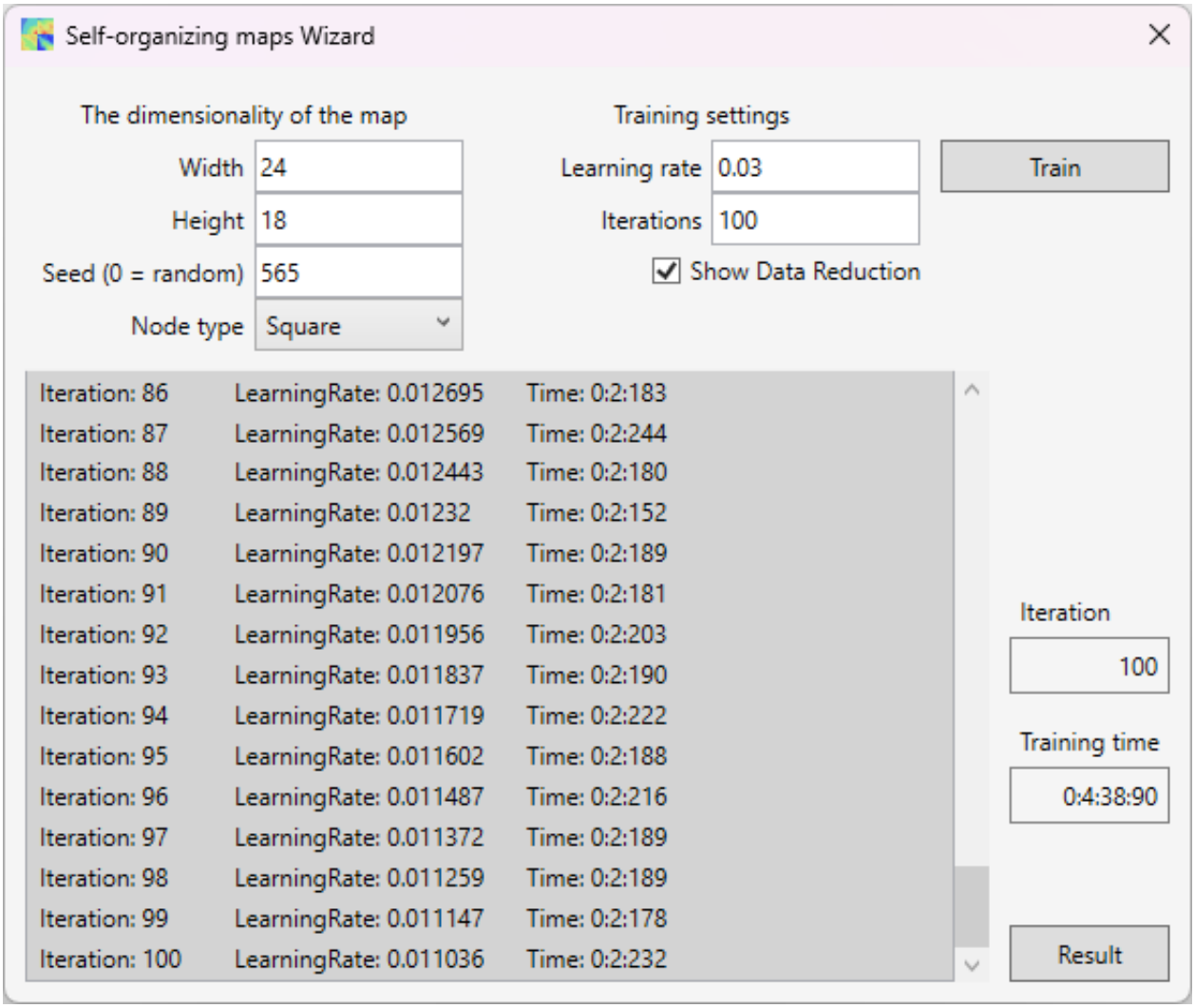


Рис. 4. Налаштування параметрів моделі SOM та запуск процесу навчання з увімкненою функцією Show Data Reduction

Після завершення первинного навчання SOM, за умови активованої опції Show Data Reduction, автоматично виконується третій етап – аналіз варіативності ознак за алгоритмом IQR. На отриманих картах результатів кластеризації (рис. 5) візуально позначено ознаки, які були визначені як малозначущі відповідно до міжквартильного критерію. Такі ознаки відображаються червоним кольором і супроводжуються позначкою Low Influence, що сигналізує про їхній низький внесок у формування кластерів. До цієї групи потрапили, зокрема, ознаки AverageConversationDuration, OtherCountriesCalls та ShareOfLandlineCalls (табл. 1). Це свідчить про те, що їх варіативність у структурі даних є незначною, і тому вони можуть бути виключені без суттєвої втрати інформаційного змісту всієї сукупності ознак.

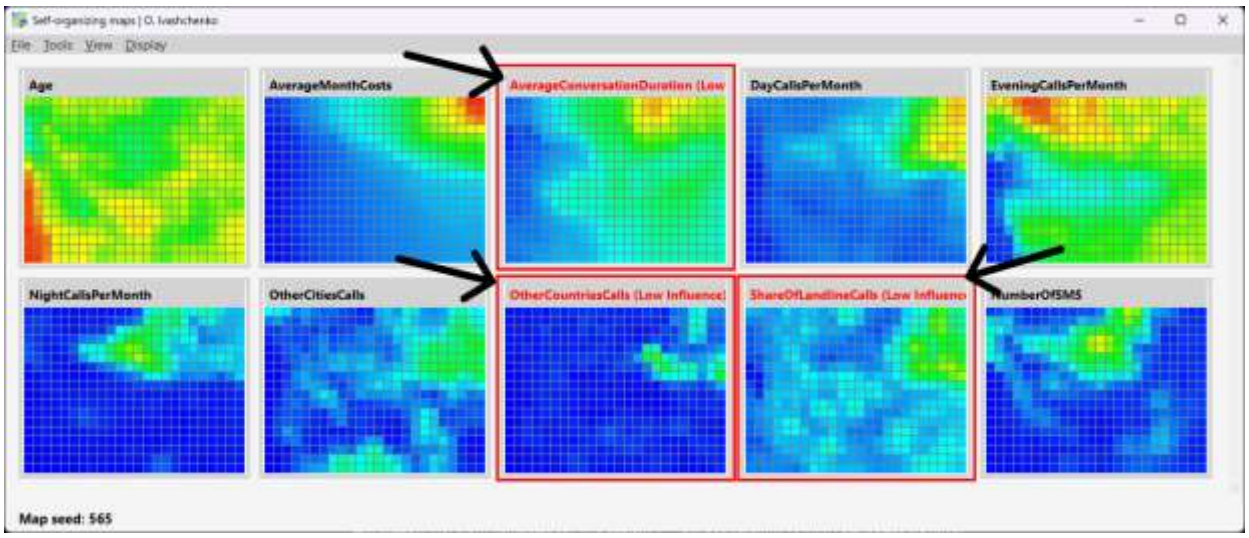


Рис. 5. Результати кластеризації із позначенням ознак, визначених алгоритмом як малозначущі (Low Influence)

На четвертому етапі з набору вхідних даних вилучаються ознаки, визначені за результатами попереднього аналізу як малозначущі, що дає змогу сформувати скорочений набір ознак для подальшої обробки. На п’ятому етапі здійснюється повторне навчання самоорганізаційної карти Кохонена (рис. 6) на оновленому наборі даних із використанням параметрів моделі, визначених під час другого етапу первинного навчання (рис. 4), але без застосування процедури редукції ознак.

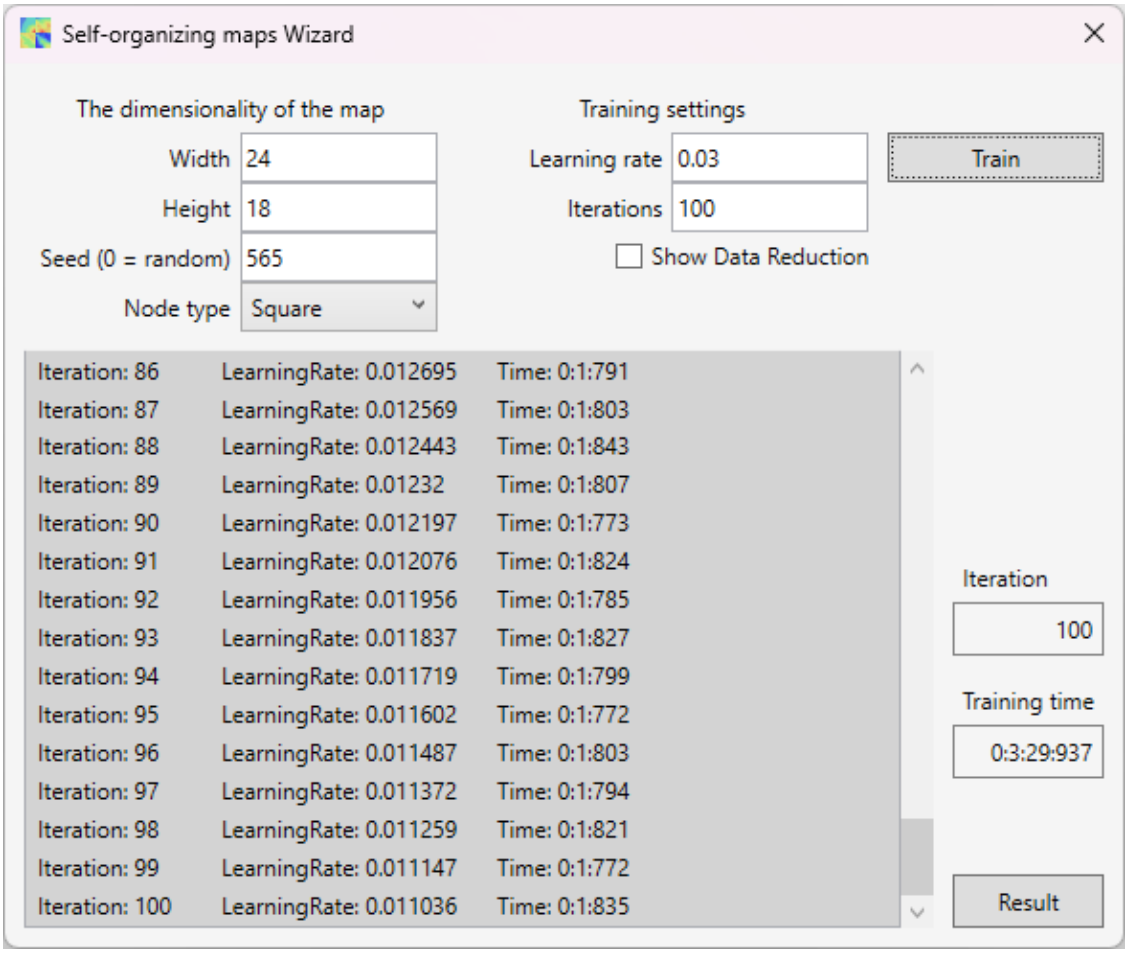


Рис. 6. Запуск повторного навчання SOM після виключення малозначущих ознак з навчальної вибірки

Отримані результати повторної кластеризації (рис. 7) демонструють формування карти Кохонена на основі скороченого набору даних, що дає змогу провести шостий етап – порівняння результатів кластеризації після редукції ознак із попереднім варіантом навчання, у якому використовувалася повна вибірка.

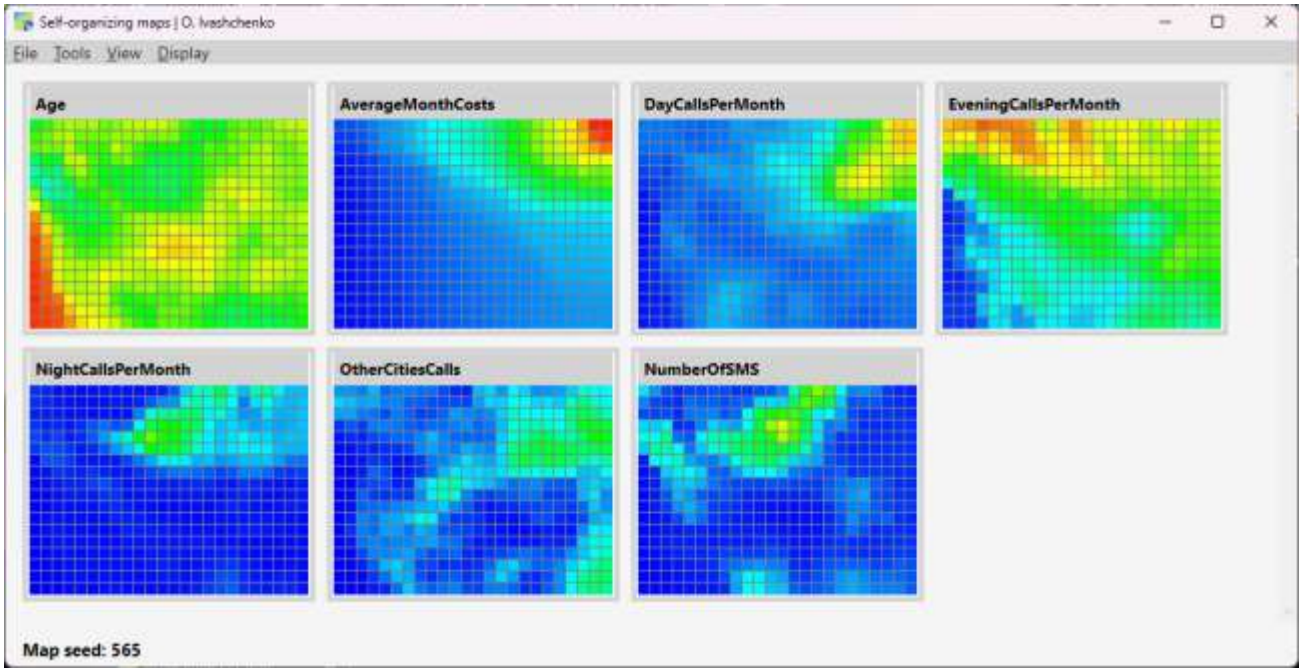


Рис. 7. Результати кластеризації після виключення малозначущих ознак

Результати дослідження

У результаті проведеного дослідження побудовано дві моделі самоорганізаційних карт Кохонена: перша – на основі всіх ознак навчальної вибірки, і друга – після застосування методу редукції даних. Навчання моделей здійснювалося з використанням однакових параметрів, що забезпечує коректність порівняння отриманих результатів. Зокрема, незмінними залишалися такі параметри: розмір карти (24×18), кількість епох навчання (100), швидкість навчання та початкове значення параметра *seed* = 565, який використовується для ініціалізації генератора випадкових чисел. Завдяки однаковому значенню *seed* результати кластеризації можна відтворити, що забезпечує збіжність експерименту при порівнянні двох моделей.

Отримані результати показали, що карти, побудовані до та після редукції даних, мають дуже подібну топологію. Візуальний аналіз продемонстрував, що сформовані кластери на обох картах розташовані майже ідентично, а загальна структура розподілу даних зберігається. Незначні відмінності спостерігаються лише в межових ділянках між окремими групами, що свідчить про те, що вилучення малозначущих ознак не призвело до суттєвих змін у структурі моделі.

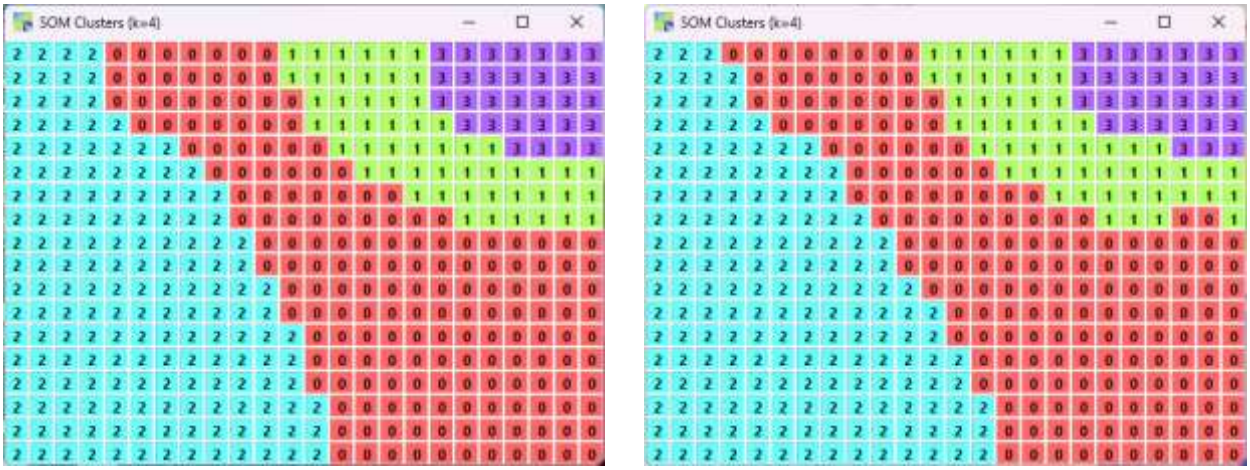
Також час навчання SOM відрізнявся для моделей до та після редукції даних: після редукції спостерігалося зменшення тривалості навчання, що зумовлено скороченням кількості вхідних ознак і відповідним зниженням обчислювальних витрат.

Для підтвердження схожості результатів кластеризації карту SOM додатково сегментовано за допомогою алгоритму k-means, який на основі позицій нейронів SOM виконує поділ карти на задану кількість кластерів за критерієм мінімальної відстані між векторами вагових коефіцієнтів нейронів і центрами кластерів за Евклідовою метрикою.

З метою перевірки стабільності моделі кластеризацію проведено для двох

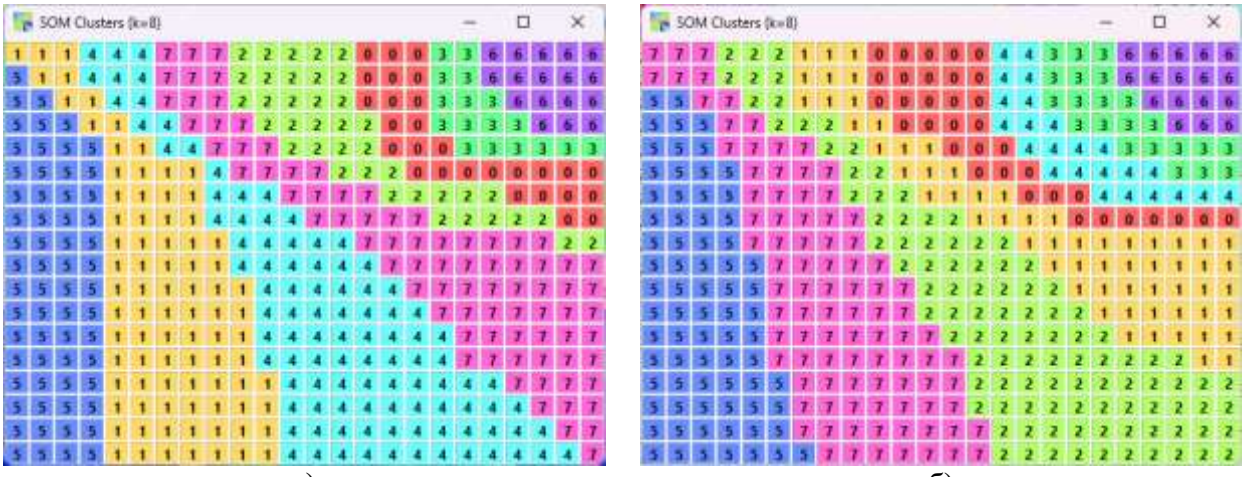
значень параметра *k*: 4 та 8. Такий метод дозволяє оцінити, наскільки зберігається топологічна структура карти при різних рівнях деталізації поділу даних.

Отримані результати кластеризації (рис. 8а, 8б – порівняння до та після редукції при *k* = 4; рис. 9а, 9б – порівняння до та після редукції при *k* = 8) свідчать про практично однаковий просторовий розподіл кластерів. Межі кластерів збігаються з мінімальними відхиленнями, що підтверджує збереження топологічної структури після усунення малозначущих ознак. Таким чином, редукція даних не вплинула негативно на здатність SOM до виявлення закономірностей, що підтверджується збереженням топологічної структури карти та збігом меж кластерів при різних значеннях параметра *k*.



а) б)

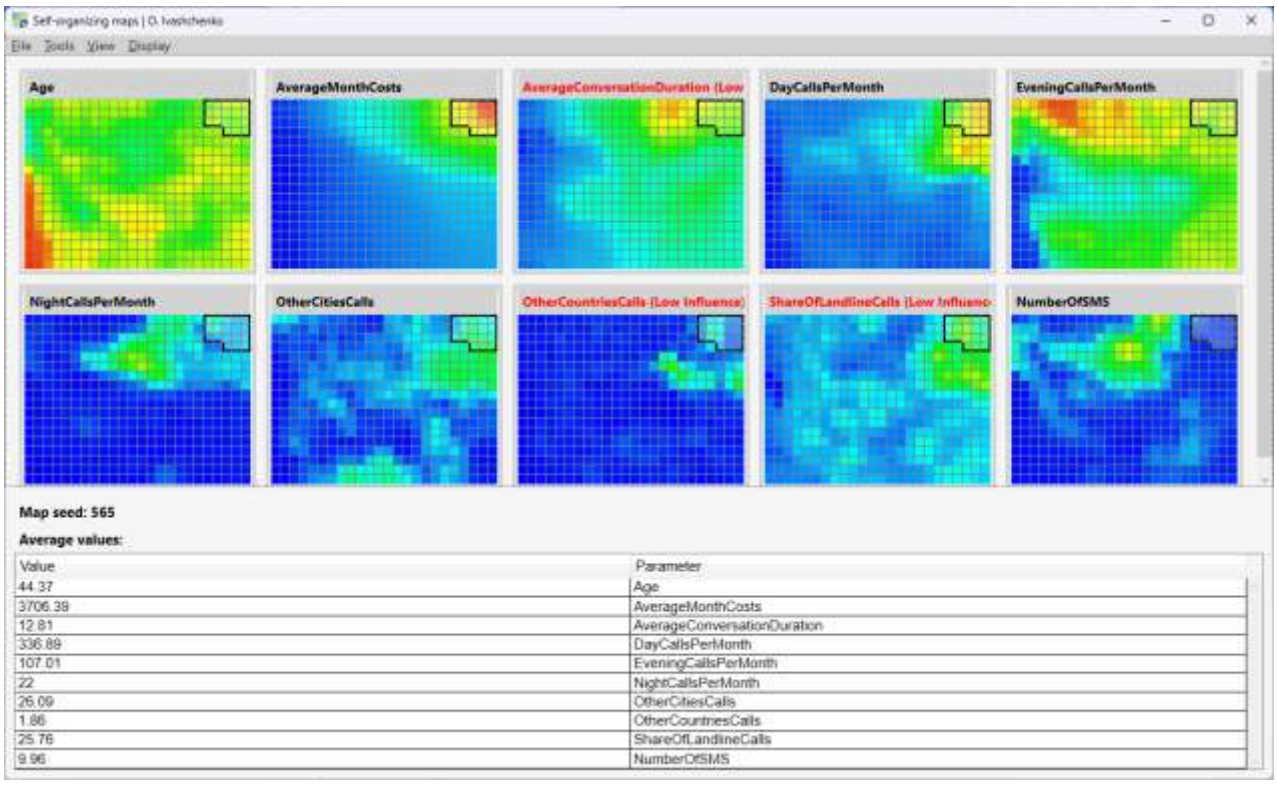
Рис. 8. Порівняння результатів кластеризації карти SOM до та після редукції ознак при параметрі *k* = 4: до редукції (а); після редукції (б)



а) б)

Рис. 9. Порівняння результатів кластеризації карти SOM до та після редукції ознак при параметрі *k* = 8: до редукції (а); після редукції (б)

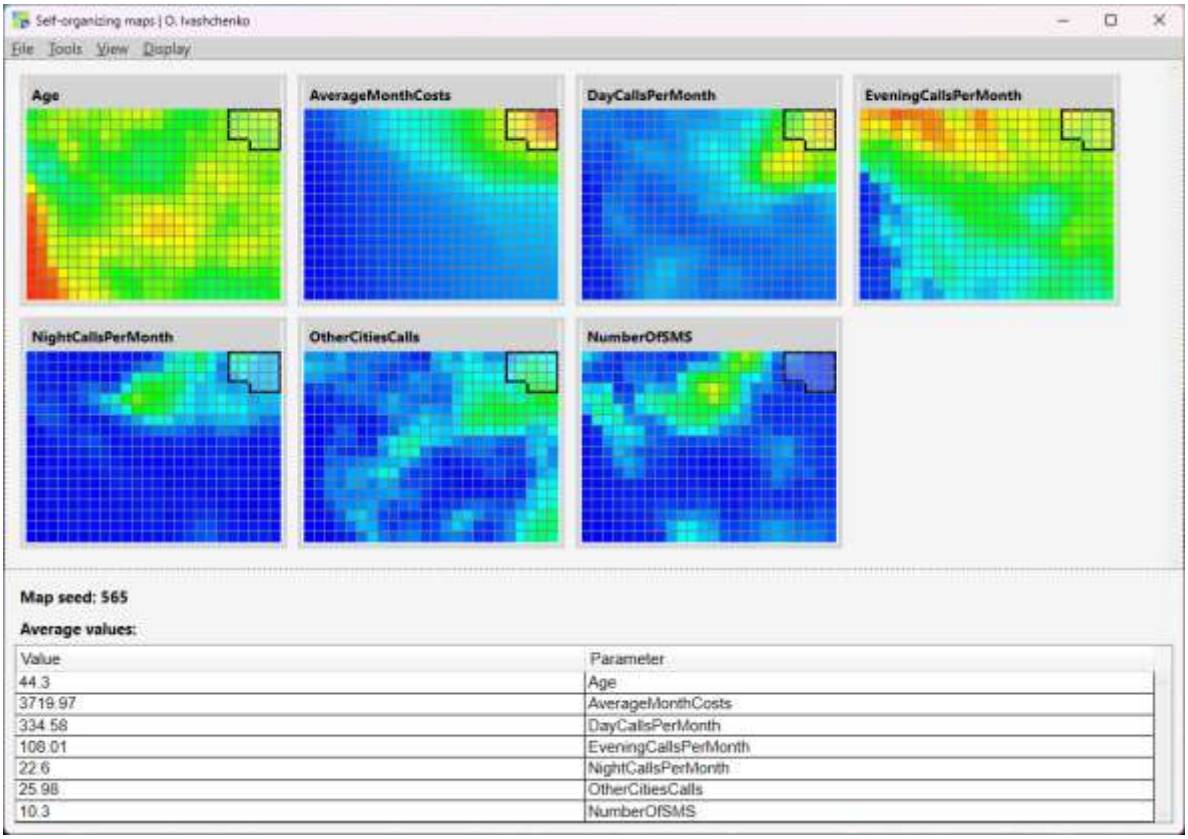
Для кількісного оцінювання відмінностей між результатами кластеризації до та після редукції даних проведено порівняльний аналіз середніх значень ознак у межах однакових ділянок карти Кохонена. Для цього на побудованій карті обрано групу вузлів, що утворюють локальний кластер (рис. 10), після чого в програмному додатку автоматично обчислено середні значення для всіх ознак спостережень, віднесених до цих вузлів.



Value	Parameter
44.37	Age
3706.39	AverageMonthCosts
12.81	AverageConversationDuration
336.89	DayCallsPerMonth
107.01	EveningCallsPerMonth
22	NightCallsPerMonth
26.09	OtherCitiesCalls
1.86	OtherCountriesCalls
25.76	ShareOfLandlineCalls
9.96	NumberOfSMS

Рис. 10. Вибір вузлів карти Кохонена для обчислення середніх значень ознак (до редукції даних)

Аналогічну процедуру виконано для карти, отриманої після редукції даних (рис. 11).



Value	Parameter
44.3	Age
3719.97	AverageMonthCosts
334.58	DayCallsPerMonth
108.01	EveningCallsPerMonth
22.6	NightCallsPerMonth
25.98	OtherCitiesCalls
10.3	NumberOfSMS

Рис. 11. Вибір вузлів карти Кохонена для обчислення середніх значень ознак (після редукції даних)

Оцінювання подібності результатів кластеризації до та після редукції даних виконано на основі статистичних критеріїв, що характеризують абсолютні та відносні відхилення між середніми значеннями ознак (табл. 2).

Таблиця 2. Середні значення ознак до та після редукції даних

Ознака	До редукції	Після редукції
Age	44,37	44,3
AverageMonthCosts	3706,39	3719,97
DayCallsPerMonth	336,89	334,58
EveningCallsPerMonth	107,01	108,01
NightCallsPerMonth	22	22,6
OtherCitiesCalls	26,09	25,98
NumberOfSMS	9,96	10,3

Зокрема, використано середнє абсолютне відхилення (*MAD* – Mean Absolute Deviation) (1), середню квадратичну похибку (*MSE* – Mean Squared Error) (2), корінь середньої квадратичної похибки (*RMSE* – Root Mean Squared Error) (3), середню відносну похибку (*MPE* – Mean Percentage Error) (4), середню абсолютну відносну похибку (*MAPE* – Mean Absolute Percentage Error) (5) [26] та показник схожості (6), який відображає узагальнену оцінку подібності результатів.

Додатково для кількісного аналізу кореляційного зв'язку між наборами даних обчислено коефіцієнт детермінації *D* (7) та коефіцієнт кореляції *R* (8), які характеризують ступінь взаємозалежності між значеннями ознак до та після редукції [27].

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|; \quad (1)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2; \quad (2)$$

де Y_i – середнє значення ознаки до редукції, $\hat{Y}_i$ – середнє значення після редукції, $\bar{Y}$ - сукупне середнє значення всіх ознак, n – кількість ознак.

$$RMSE = \sqrt{MSE}; \quad (3)$$

$$MPE = \frac{100}{n}\sum_{i=1}^{n}\frac{Y_i - \hat{Y}_i}{Y_i}; \quad (4)$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|; \quad (5)$$

$$Similarity = 100 - MAPE; \quad (6)$$

$$D = \frac{\left[n\cdot\sum_{i=1}^{n}(Y_i\hat{Y}_i) - \sum_{i=1}^{n}Y_i\cdot\sum_{i=1}^{n}\hat{Y}_i\right]^2}{\left[n\cdot\sum_{i=1}^{n}{Y_i}^2 - (\sum_{i=1}^{n}Y_i)^2\right]\left[n\cdot\sum_{i=1}^{n}{\hat{Y}_i}^2 - (\sum_{i=1}^{n}\hat{Y}_i)^2\right]}; \quad (7)$$

$$R = \sqrt{D}, \quad (8)$$

Результати розрахунків за критеріями (1)-(8) наведено у таблиці 3.

Таблиця 3. Результати розрахунку статистичних критеріїв

MAD	MSE	RMSE	MPE	MAPE	Similarity	D	R
2,573	27,321	5,227	-0,882 %	1,244 %	98,756 %	0,9996	0,9998

Показники, наведені в таблиці 3, характеризують відхилення між середніми значеннями ознак до та після редукції, а також відображають рівень подібності й кореляційної узгодженості між обома наборами даних (табл. 2).

З використанням системи статистичного аналізу STATISTICA перевірено гіпотезу щодо значущості відмінностей між середніми значеннями ознак до та після редукції даних із використанням *t*-критерію для залежних вибірок при рівні значущості p=0,05.

T-test for Dependent Samples (Spreadsheet1)
Marked differences are significant at p < .05000

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95.000%	Confidence +95.000%
Before DR	607.530	1371.244								
After DR	609.391	1376.320	7	-1.861	5.276	-0.934	6	0.387	-6.741	3.018

Рис. 12. Оцінка значущої відмінності між середніми значеннями ознак до та після редукції даних за t-критерієм

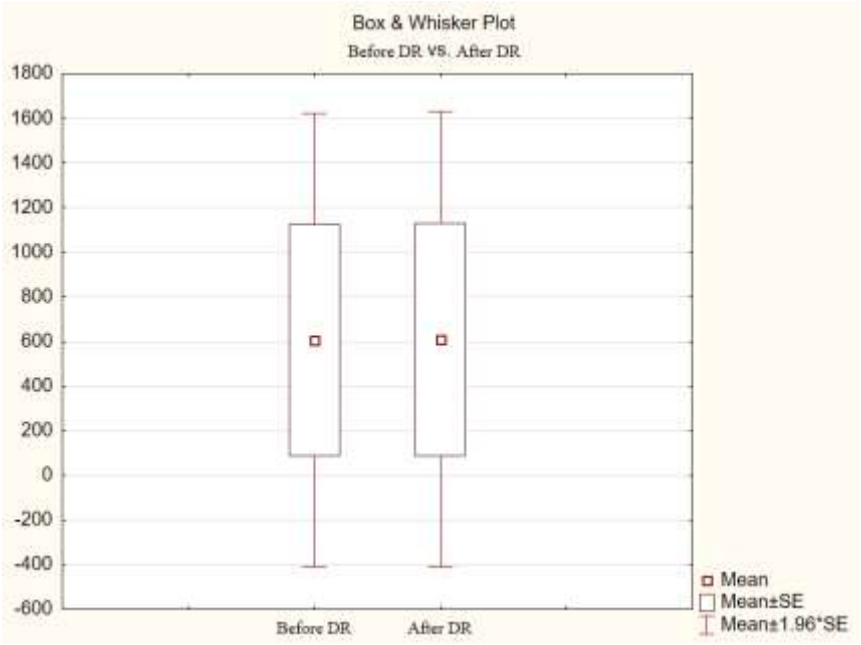


Рис. 13. Графік розмаху для порівняння розподілу значень до та після редукції даних

Отримане значення p=0,387 перевищує пороговий рівень значущості (p > 0,05), що свідчить про відсутність статистично значущих відмінностей між вибірками (рис. 12).

Як видно з графіка (рис. 13), пороги стандартних похибок свідчать про відсутність значущої відмінності у структурі даних до та після редукції.

Таким чином, встановлено, що редукція даних не вплинула на середні значення кожної ознаки у межах обраних вузлів на карті, а результати кластеризації, які отримано методом SOM, можна

вважати стабільними та достовірними. Застосування редукції даних дозволило скоротити кількість вхідних ознак без втрати інформативності та зберегти високу подібність результатів кластеризації.

Обговорення результатів

Отримані результати дослідження підтверджують, що попередня редукція вхідних ознак є доцільним етапом під час побудови самоорганізаційних карт Кохонена. Запропонований метод редукції даних на основі алгоритму IQR дозволяє автоматично визначати ознаки з низькою варіативністю, що дає змогу усунути надлишкові та малозначущі ознаки й підвищити стабільність результатів кластеризації.

Редукція даних сприяла зменшенню обчислювальних витрат і скороченню часу навчання моделі: до редукції час становив 4 хв 38 с (рис. 4), тоді як після редукції – 3 хв 29 с (рис. 6), що відповідає скороченню тривалості навчання на 69 секунд для вибірки обсягом 1000 спостережень. Навіть за такого обсягу даних отриманий результат демонструє тенденцію до підвищення ефективності навчання SOM після виключення малозначущих ознак. У разі масштабування задачі на значно більші вибірки, характерні для Big Data, очікуваний виграш у часі зростатиме, що робить застосування методу редукції особливо доцільним і може мати практичне значення для підвищення продуктивності систем обробки даних.

Порівняння карт SOM, побудованих до і після редукції даних, показало їх високу подібність за топологічною структурою – візуально межі кластерів практично збігаються. Також варто зазначити, що під час використання алгоритму k-means із фіксованими значеннями $k = 4$ та $k = 8$ для карт SOM до та після редукції даних (рис. 8, рис. 9) розподіл кластерів залишився практично незмінним після редукції даних. Це підтверджує, що SOM зберігає стійкість до зменшення розмірності вхідних даних за умови коректного визначення малозначущих ознак.

Порівняння середніх значень ознак до та після редукції демонструє, що зміни є незначними і не впливають на загальну структуру даних. Для більшості показників коливання знаходяться в межах десятих часток одиниці, що в контексті їх абсолютних величин свідчить про стабільність статистичного профілю вибірки (табл. 2). Це підтверджується значенням середнього абсолютного відхилення *MAD* = 2,573 (табл. 3), яке є малим порівняно з діапазонами відповідних ознак, що вимірюються десятками чи сотнями одиниць. Низькі значення *MSE* та *RMSE* (27,321 і 5,227 відповідно) також вказують на те, що сумарна квадратична різниця між двома наборами даних залишається мінімальною та не має суттєвого впливу на форму розподілу ознак. Відносні показники *MPE* та *MAPE* демонструють, що редукція призвела лише до незначних відхилень на рівні менш ніж (1,0–1,3) %, що не перевищує звичайний рівень коливань значень у вибірці. Отримане значення критерію схожості *Similarity* (98,756 %) свідчить про майже повний збіг профілів ознак до та після редукції. Це означає, що взаємозв’язки між ознаками, їхня пропорційність та структура розподілу залишилися практично незмінними.

У сукупності наведені результати дозволяють зробити висновок, що процедура редукції ознак не порушує статистичної цілісності вибірки та зберігає її ключові характеристичні властивості. Застосовані критерії однозначно підтверджують, що редукція не вплинула на базову структуру даних і може розглядатися як коректний етап підготовки вхідного простору для SOM.

Для кількісної оцінки взаємозв’язку між середніми значеннями ознак до та після редукції даних розраховано коефіцієнт детермінації D = 0,9996 та коефіцієнт кореляції R = 0,9998. Отримані результати свідчать про наявність майже повного функціонального зв’язку між наборами даних, що вказує на практично ідеальну подібність результатів кластеризації після редукції. Високе значення коефіцієнта R означає, що редукція не змінила структуру

розподілу ознак і зберегла внутрішні залежності між ними, а коефіцієнт *D* підтверджує, що понад 99,9 % варіації початкових даних пояснюється моделлю після редукції.

Для якісної інтерпретації рівня зв'язку між змінними використано шкалу Чеддока (Chaddock scale) [28], що подає відповідність між кількісними та якісними оцінками коефіцієнта кореляції (табл. 4). Відповідно до наведеної шкали, значення $R = 0{,}9998$ належить до діапазону $0{,}99 < R < 1$, який характеризується як функціональний зв'язок. Це підтверджує, що після редукції даних SOM відтворює результати кластеризації з максимальною точністю та без суттєвих втрат інформації.

Таблиця 4. Відповідність між кількісними та якісними оцінками коефіцієнта кореляції за шкалою Чеддока [28]

Кількісна характеристика зв'язку	**Якісна характеристика зв'язку**
0<R<0,1	Відсутній зв'язок
0,1<R<0,3	Слабкий зв'язок
0,3 <R<0,5	Помірний зв'язок
0,5<R<0,7	Відчутний зв'язок
0,7<R<0,9	Тісний зв'язок
0,9<R<0,99	Сильний зв'язок
0,99<R<1	Функціональний зв'язок

Таким чином, встановлено, що попередня редукція даних сприяє покращенню інтерпретованості результатів, зберігаючи при цьому їхню точність і відтворюваність.

Запропонований метод можна розглядати як раціональне рішення для задач кластерного аналізу у сферах, де наявні великі масиви даних. Інтеграція методу редукції даних безпосередньо у програмне середовище побудови SOM підвищує рівень автоматизації процесу аналізу та розширює практичні можливості цього інструмента. Завдяки цьому дослідник або аналітик отримує можливість не лише формувати кластери, але й оцінювати якість кожної ознаки з точки зору її впливу на результат, що підсилює аналітичну цінність отриманих карт Кохонена.

Результати роботи доводять ефективність інтеграції методу редукції даних у процес побудови самоорганізаційних карт і відкривають перспективи для подальших досліджень, пов'язаних з автоматичним налаштуванням порогів відсіву та адаптивним вибором методів статистичної оцінки інформативності ознак.

Висновки

1. Розроблено метод редукції даних, заснований на статистичному аналізі варіативності ознак з використанням алгоритму міжквартильного розмаху (*IQR*). Показано, що запропонований метод забезпечує автоматичне виявлення малозначущих ознак і їх вилучення до етапу навчання самоорганізаційної карти Кохонена.

2. Проведено експериментальне дослідження для перевірки ефективності запропонованого методу на прикладі кластеризації даних до та після редукції з використанням однакової вибірки та ідентичних параметрів навчання, включно з фіксованим значення *seed* = 565. Результати показали збереження топологічної структури карти SOM та ідентичність кластерного розподілу при $k = 4$ і $k = 8$. Аналіз статистичних критеріїв засвідчив мінімальні відхилення між середніми значеннями ознак до та після редукції (*MAD* = 2,573; *MSE* = 27,321; *RMSE* = 5,227), а відносні показники *MPE* та *MAPE* підтвердили, що зміни залишаються в межах статистичної мінливості вибірки. Високе значення критеріїв *Similarity* (98,756 %), $D = 0{,}9996$ та кореляції $R = 0{,}9998$ свідчить про практично повну відповідність між наборами даних, що

підтверджує збереження статистичної цілісності даних після редукції. Після редукції час навчання SOM скоротився з 4 хв 38 с до 3 хв 29 с, що для вибірки обсягом 1000 спостережень відповідає зменшенню тривалості обчислень приблизно на 24,8 %, що свідчить про зменшення обчислювальних витрат без втрати точності та стабільності кластеризації. Таким чином, запропонований метод редукції даних підвищує ефективність та інтерпретованість результатів кластеризації, зберігаючи їх точність, стабільність і достовірність при зменшенні обчислювальних витрат.

Література

1. Dimensionality Reduction using Self Organizing Maps. [Online]. Available:https://stackoverflow.com/questions/27086222/dimensionality-reduction-using-self-organizing-maps/
2. Zabielin S. Big Data Analysis via Model Reduction Methods // System Research and Information Technologies. – 2018. – № 2. – pp. 35-41. doi:https://doi.org/10.20535/SRIT.2308-8893.2018.2.04
3. Vayssieres M. Master Kohonen Self-Organizing Maps: A Hands-On Guide to Data Exploration with Python. [Online]. Available:https://medium.com/@MahounaVAYSSIERES/master-kohonen-self-organizing-maps-a-hands-on-guide-to-data-exploration-with-python-fb92f8ebd6f6
4. Agboka K. M., Abdel-Rahman E. M., Salifu D., Kanji B., Ndjomatchoua F. T., Guimapi R. A. Y., Ekesi S., Landmann T. Towards combining self-organizing maps (SOM) and convolutional neural network (CNN) for improving model accuracy: Application to malaria vectors phenotypic resistance // MethodsX. – 2025. – Vol. 14. – 103198. doi: https://doi.org/10.1016/j.mex.2025.103198
5. Salap-Ayca S. Self-organizing maps as a dimension reduction approach for spatial global sensitivity analysis visualization // Transactions in GIS. – 2025. – Vol. 26. – № 4. – pp. 1718-1734. doi: https://doi.org/10.1111/tgis.12963
6. Silva R. G., Wilcox S. J. Feature evaluation and selection for condition monitoring using a self-organizing map and spatial statistics // Artificial Intelligence for Engineering Design, Analysis and Manufacturing. – 2018. – Vol. 33. – pp. 1-10. doi: 10.1017/S0890060417000518.
7. Khacef L., Rodriguez L., Miramond B. Improving Self-Organizing Maps with Unsupervised Feature Extraction // The International Conference on Neural Information Processing. – 2020. doi: https://doi.org/10.48550/arXiv.2009.02174
8. Fan X., Zhang S., Xue X., Jiang R., Fan S., Kou H. An Improved Self-Organizing Map (SOM) Based on Virtual Winning Neurons // Symmetry. – 2025. – Vol. 17. – № 3. – pp. 449. doi: 10.3390/sym17030449
9. Starkey A., Akpan U. I., Al Hosni O., Pullissery Y. Class-Level Feature Selection Method Using Feature Weighted Growing Self-Organising Maps // arXiv Preprint. – 2025. doi: https://doi.org/10.48550/arXiv.2503.11732
10. Benabdeslem K., Lebbah M. Feature Selection for Self-Organizing Map // 29th International Conference on Information Technology Interfaces. – Cavtat, Croatia. – 2007. – pp. 45-50. doi: https://doi.org/10.1109/ITI.2007.4283742
11. Ceylan O., Taskin Kaya G. Feature Selection Using Self Organizing Map Oriented Evolutionary Approach // IEEE International Geoscience and Remote Sensing Symposium (IGARSS). – Brussels, Belgium. – 2021. – pp. 4003-4006. doi:https://doi.org/10.1109/IGARSS47720.2021.9553491
12. Pratiwi D. The Use of Self Organizing Map Method and Feature Selection in Image Database Classification System // International Journal of Computer Science. – 2012. – № 9.
13. Khelil M. I., Ladjal M., Brik Y., Ouali M. A. Self-Organizing Maps-Based Features Selection with Deep LSTM and SVM Classification Approaches for Advanced Water Quality Monitoring // International Journal of Intelligent Engineering and Systems. – 2022. – Vol. 15. – № 3. – pp. 90-102. doi: https://doi.org/10.22266/ijies2022.0630.09
14. Zang Z., Xu Y., Lu L., Geng Y., Yang S., Li S. Z. UDRN: Unified Dimensional Reduction Neural Network for feature selection and feature projection // Neural Networks. – 2023. – Vol. 161. – pp. 626-637. doi: https://doi.org/10.1016/j.neunet.2023.02.018
15. Mwangi B., Tian S. T., Jair C. S. A Review of Feature Reduction Techniques in Neuroimaging // Neuroinformatics. – 2014. – Vol. 12. – № 2. – pp. 229-244. doi: https://doi.org/10.1007/s12021-013-9204-3
16. Cheng X. A Comprehensive Study of Feature Selection Techniques in Machine Learning Models // Insights in Computer, Signals and Systems. – 2024. – Vol. 1. – pp. 65-78. doi: https://doi.org/10.70088/xpf2b276
17. Rossi R., Murari A., Gelfusa M. A deep learning framework for feature selection and dimensional analysis: Variational explainable neural networks // Knowledge-Based Systems. – 2025. – Vol. 324. – 113940. doi: https://doi.org/10.1016/j.knosys.2025.113940
18. Passimier A., Folco P., Raimondi D., Birolo G., Moreau Y., Fariselli P. A quantitative benchmark of neural network feature selection methods onomics data // Scientific Reports. – 2024. – Vol. 14. – 31180. doi: https://doi.org/10.1038/s41598-024-82583-5
19. Trelina A., Procházka A. Binary Stochastic Filtering: a Solution for Supervised Feature Selection and Neural Network Shape Optimization // arXiv Preprint. – 2019. doi: https://doi.org/10.48550/arXiv.1902.04510

20. Ougiaroglou S., Diamantaras K. I., Evangelidis G. Exploring the effect of data reduction on neural network and support vector machine classification // Neurocomputing. – 2018. – Vol. 280. – pp. 101-110. doi: https://doi.org/10.1016/j.neucom.2017.08.076

21. Jia W., Sun M., Liang J., Hou S. Feature dimensionality reduction: a review // Complex & Intelligent Systems. – 2022. – Vol. 8. – pp. 2663-2693. doi: https://doi.org/10.1007/s40747-021-00650-0

22. Ivashchenko O., Fedin S. Improving The Som Algorithm To Ensure Stability And Reproducibility Of Data Clustering Results // System Research and Information Technologies. – 2025. – № 4.

23. Іващенко О. В., Федін С. С. Оптимізація алгоритму Кохонена для забезпечення відтворюваності результатів кластеризації // Штучний інтелект та інформаційні технології: матеріали Першої міжнародної науково-практичної конференції (3–4 червня 2024 р., Київ, Україна). – Київ: НУХТ. – 2024. – pp. 226-227.

24. Bhandari P. How to Find Interquartile Range (IQR) | Calculator & Examples. [Online]. Available: https://www.scribbr.com/statistics/interquartile-range/

25. Ivashchenko O., Fedin S. Supporting Decision-Making in the Segmentation of Telecommunications Company Customers Using Specialized Software // Modern Engineering and Innovative Technologies. – 2025. – № 41-01. – pp. 116-136. doi:https://doi.org/10.30890/2567-5273.2025-41-01-002

26. Jonatasv. Metrics Evaluation: MSE, RMSE, MPE and MAPE. – 2024. [Online]. Available: https://medium.com/@jonatasv/metrics-evaluation-mse-rmse-mae-and-mape-317cab85a26b

27. Shaun Turney. Coefficient of Determination (R^2) | Calculation & Interpretation. – 2023. [Online]. Available:https://www.scribbr.com/statistics/coefficient-of-determination/

28. Ugur Turan. A Correlation Coefficients Analysis on Innovative Sustainable Development Groups // EUREKA: Social and Humanities. – 2020. – Vol. 1. – № 1. – pp. 46-55. doi: https://doi.org/10.21303/2504-5571.2020.001130

References

1. Dimensionality Reduction using Self Organizing Maps. [Online]. Available:https://stackoverflow.com/questions/27086222/dimensionality-reduction-using-self-organizing-maps/

2. Zabielin S. Big Data Analysis via Model Reduction Methods // System Research and Information Technologies. – 2018. – № 2. – pp. 35-41. https://doi.org/10.20535/SRIT.2308-8893.2018.2.04

3. Vayssieres M. Master Kohonen Self-Organizing Maps: A Hands-On Guide to Data Exploration with Python. [Online]. Available:https://medium.com/@MahounaVAYSSIERES/master-kohonen-self-organizing-maps-a-hands-on-guide-to-data-exploration-with-python-fb92f8ebd6f6

4. Agboka K. M., Abdel-Rahman E. M., Salifu D., Kanji B., Ndjomatchoua F. T., Guimapi R. A. Y., Ekesi S., Landmann T. Towards combining self-organizing maps (SOM) and convolutional neural network (CNN) for improving model accuracy: Application to malaria vectors phenotypic resistance // MethodsX. – 2025. – Vol. 14. – 103198. https://doi.org/10.1016/j.mex.2025.103198

5. Salap-Ayca S. Self-organizing maps as a dimension reduction approach for spatial global sensitivity analysis visualization // Transactions in GIS. – 2025. – Vol. 26. – № 4. – pp. 1718-1734. https://doi.org/10.1111/tgis.12963

6. Silva R. G., Wilcox S. J. Feature evaluation and selection for condition monitoring using a self-organizing map and spatial statistics // Artificial Intelligence for Engineering Design, Analysis and Manufacturing. – 2018. – Vol. 33. – pp. 1-10. doi: 10.1017/S0890060417000518.

7. Khacef L., Rodriguez L., Miramond B. Improving Self-Organizing Maps with Unsupervised Feature Extraction // The International Conference on Neural Information Processing. – 2020. https://doi.org/10.48550/arXiv.2009.02174

8. Fan X., Zhang S., Xue X., Jiang R., Fan S., Kou H. An Improved Self-Organizing Map (SOM) Based on Virtual Winning Neurons // Symmetry. – 2025. – Vol. 17. – № 3. – pp. 449. doi: 10.3390/sym17030449

9. Starkey A., Akpan U. I., Al Hosni O., Pullissery Y. Class-Level Feature Selection Method Using Feature Weighted Growing Self-Organising Maps // arXiv Preprint. – 2025. https://doi.org/10.48550/arXiv.2503.11732

10. Benabdeslem K., Lebbah M. Feature Selection for Self-Organizing Map // 29th International Conference on Information Technology Interfaces. – Cavtat, Croatia. – 2007. – pp. 45-50. https://doi.org/10.1109/ITI.2007.4283742

11. Ceylan O., Taskin Kaya G. Feature Selection Using Self Organizing Map Oriented Evolutionary Approach // IEEE International Geoscience and Remote Sensing Symposium (IGARSS). – Brussels, Belgium. – 2021. – pp. 4003-4006. https://doi.org/10.1109/IGARSS47720.2021.9553491

12. Pratiwi D. The Use of Self Organizing Map Method and Feature Selection in Image Database Classification System // International Journal of Computer Science. – 2012. – № 9.

13. Khelil M. I., Ladjal M., Brik Y., Ouali M. A. Self-Organizing Maps-Based Features Selection with Deep LSTM and SVM Classification Approaches for Advanced Water Quality Monitoring // International Journal of Intelligent Engineering and Systems. – 2022. – Vol. 15. – № 3. – pp. 90-102. https://doi.org/10.22266/ijies2022.0630.09

14. Zang Z., Xu Y., Lu L., Geng Y., Yang S., Li S. Z. UDRN: Unified Dimensional Reduction Neural Network for feature selection and feature projection // Neural Networks. – 2023. – Vol. 161. – pp. 626-637. https://doi.org/10.1016/j.neunet.2023.02.018

15. Mwangi B., Tian S. T., Jair C. S. A Review of Feature Reduction Techniques in Neuroimaging //

Neuroinformatics. – 2014. – Vol. 12. – № 2. – pp. 229-244. https://doi.org/10.1007/s12021-013-9204-3

16. Cheng X. A Comprehensive Study of Feature Selection Techniques in Machine Learning Models // Insights in Computer, Signals and Systems. – 2024. – Vol. 1. – pp. 65-78. https://doi.org/10.70088/xpf2b276

17. Rossi R., Murari A., Gelfusa M. A deep learning framework for feature selection and dimensional analysis: Variational explainable neural networks // Knowledge-Based Systems. – 2025. – Vol. 324. – 113940. https://doi.org/10.1016/j.knosys.2025.113940

18. Passimier A., Folco P., Raimondi D., Birolo G., Moreau Y., Fariselli P. A quantitative benchmark of neural network feature selection methods onomics data // Scientific Reports. – 2024. – Vol. 14. – 31180. https://doi.org/10.1038/s41598-024-82583-5

19. Trelina A., Procházka A. Binary Stochastic Filtering: a Solution for Supervised Feature Selection and Neural Network Shape Optimization // arXiv Preprint. – 2019. https://doi.org/10.48550/arXiv.1902.04510

20. Ougiaroglou S., Diamantaras K. I., Evangelidis G. Exploring the effect of data reduction on neural network and support vector machine classification // Neurocomputing. – 2018. – Vol. 280. – pp. 101-110. https://doi.org/10.1016/j.neucom.2017.08.076

21. Jia W., Sun M., Liang J., Hou S. Feature dimensionality reduction: a review // Complex & Intelligent Systems. – 2022. – Vol. 8. – pp. 2663-2693. https://doi.org/10.1007/s40747-021-00650-0

22. Ivashchenko O., Fedin S. Improving The Som Algorithm To Ensure Stability And Reproducibility Of Data Clustering Results // System Research and Information Technologies. – 2025. – № 4.

23. Ivashchenko, O. V., & Fedin, S. S. Optymizatsiia alhorytmu Kokhonena dlia zabezpechennia vidtvoriuvanosti rezultativ klasteryzatsii // Shtuchnyi intelekt ta informatsiini tekhnolohii: materialy Pershoi mizhnarodnoi naukovo-praktychnoi konferentsii (June 3–4, 2024, Kyiv, Ukraine). Kyiv: NUFT, 2024, pp. 226–227.

24. Bhandari P. How to Find Interquartile Range (IQR) | Calculator & Examples. [Online]. Available: https://www.scribbr.com/statistics/interquartile-range/

25. Ivashchenko O., Fedin S. Supporting Decision-Making in the Segmentation of Telecommunications Company Customers Using Specialized Software // Modern Engineering and Innovative Technologies. – 2025. – № 41-01. – pp. 116-136. https://doi.org/10.30890/2567-5273.2025-41-01-002

26. Jonatasv. Metrics Evaluation: MSE, RMSE, MPE and MAPE. – 2024. [Online]. Available: https://medium.com/@jonatasv/metrics-evaluation-mse-rmse-mae-and-mape-317cab85a26b

27. Shaun Turney. Coefficient of Determination (R^2) | Calculation & Interpretation. – 2023. [Online]. Available:https://www.scribbr.com/statistics/coefficient-of-determination/

28. Ugur Turan. A Correlation Coefficients Analysis on Innovative Sustainable Development Groups // EUREKA: Social and Humanities. – 2020. – Vol. 1. – № 1. – pp. 46-55. https://doi.org/10.21303/2504-5571.2020.001130

The article has been sent to the editors 14.12.25.
After processing 20.01.26.
Submitted for printing 30.03.26.

Copyright under license CCBY-SA4.0.