

УДК 004

DOI <https://doi.org/10.32689/maup.it.2023.2.1>

Наталія БАРЧЕНКО

кандидат технічних наук, доцент кафедри комп'ютерних наук, Сумський державний університет, вул. Харківська, 116, м. Суми, Україна, індекс 40007 (n.barchenko@cs.sumdu.edu.ua)

ORCID: 0000-0002-5439-8750

Володимир ЛЮБЧАК

кандидат фізико-математичних наук, доцент, завідувач кафедри кібербезпеки, Сумський державний університет, вул. Харківська, 116, м. Суми, Україна, індекс 40007 (v.liubchak@dcs.sumdu.edu.ua)

ORCID: 0000-0002-7335-6716

Дмитро ВЕЛИКОДНИЙ

кандидат фізико-математичних наук, старший викладач кафедри комп'ютерних наук, Сумський державний університет, вул. Харківська, 116, м. Суми, Україна, індекс 40007 (d.velykodnyi@cs.sumdu.edu.ua)

ORCID: 0000-0003-0044-5619

Nataliia BARCHENKO

Candidate of Technical Science, Associate Professor of the Department of Computer Science, Sumy State University, Kharkivska str., 116, Sumy, Ukraine, postal code 40007 (n.barchenko@cs.sumdu.edu.ua)

Volodymyr LIUBCHAK

Candidate of Physical and Mathematical Sciences, Associate Professor, Head of the Department of Cyber Security, Sumy State University, Kharkivska str., 116, Sumy, Ukraine, postal code 40007 (v.liubchak@dcs.sumdu.edu.ua)

Dmytro VELYKODNYI

Candidate of Physical and Mathematical Sciences, Senior Lecturer of the Department of Computer Science, Sumy State University, Kharkivska str., 116, Sumy, Ukraine, postal code 40007 (d.velykodnyi@cs.sumdu.edu.ua)

Бібліографічний опис статті: Барченко Н., Любчак В., Великодний Д. Вибір метода кластеризації з метою аналізу показників цифрових трансформацій регіонів України. *Інформаційні технології та суспільство*. 2023. Вип. 2(8). 6–17. DOI: <https://doi.org/10.32689/maup.it.2023.2.1>

Bibliographic description of the article: Barchenko N., Liubchak V., Velykodnyi D. (2023). Vybir metoda klasteryzatsii z metoiu analizu pokaznykiv tsyfrovyykh transformatsii rehioniv Ukrainy [Choice of clusterization method for digital transformation indicators analysis of regions of Ukraine]. *Informatsiini tekhnolohii ta suspilstvo – Information technology and society*, 2(8), 6–17. DOI: <https://doi.org/10.32689/maup.it.2023.2.1>

ВИБІР МЕТОДА КЛАСТЕРИЗАЦІЇ З МЕТОЮ АНАЛІЗУ ПОКАЗНИКІВ ЦИФРОВИХ ТРАНСФОРМАЦІЙ РЕГІОНІВ УКРАЇНИ¹

В статті розглянуто актуальне питання щодо застосування кластерного аналізу для опису, обґрунтування трендів та показників узгодженості розвитку регіонів України в напрямі цифрових трансформацій. Для планів розбудови та визначення тенденцій потребується систематизація наявних знань щодо трендів та інструментів цифровізації регіональних систем. Для оцінки стану цифрових трансформацій регіонів потрібно залучати показники різного типу різних напрямів діяльності, як кількісні так і якісні, що формують базу багатовимірних даних для дослідження. Метою даної роботи є дослідження та застосування кластерного аналізу для опису, обґрунтування трендів та показників узгодженості розвитку регіонів України в напрямі цифрових трансформацій. Кластерний аналіз полягає у здійсненні класифікації об'єктів дослідження за допомогою численних обчислювальних процедур. Він дає можливість класифікувати об'єкти за декількома ознаками одночасно, перетворити великий обсяг різнобічної інформації в упорядкований, компактний вигляд. Розглянуто ієрархічний метод кластеризації та метод к-середніх. Проведені експерименти в SPSS Statistics дозволили обрати в якості метода для подальших досліджень метод к-середніх. Комп'ютерний експеримент з побудови системи кластерів проведено з використанням даних звіту Міністерства цифрової трансформації України – підсумкових показників (субіндексів) цифрової трансформації регіонів. Виділено п'ять кластерів, які можна поставити у відповідність рівням розвитку цифровізації областей. Порівняння цих результатів з інтегральними індексами цифрової трансформації регіонів, які визначені за іншою методикою, демонструє гарну узгодженість. Метод к-середніх планується обрати базовим для програмної реалізації підсистеми кластеризації системи підтримки прийняття рішень щодо аналізу цифрової трансформації регіонів.

Ключові слова: кластерний аналіз, цифровізація регіонів, SPSS.

¹ Публікація підготовлена у рамках виконання наукового проєкту 0122U001232 «Реструктуризація національної економіки в напрямі цифрових трансформацій для сталого розвитку», який фінансується Національним фондом досліджень України

CHOICE OF CLUSTERIZATION METHOD FOR DIGITAL TRANSFORMATION INDICATORS ANALYSIS OF REGIONS OF UKRAINE ²

The article deals with the topical issue of the application of cluster analysis for the description and substantiation of trends and indicators of consistency of the development of the regions of Ukraine in the direction of digital transformations. Systematization of available knowledge regarding trends and tools of digitization of regional systems is required for development plans and determination of trends. In order to assess the state of digital transformations of the regions, it is necessary to attract indicators of various types of different areas of activity, both quantitative and qualitative, forming a multidimensional data base for research. The purpose of this work is the research and application of cluster analysis for the description and substantiation of trends and indicators of consistency of the development of the regions of Ukraine in the direction of digital transformations. Cluster analysis consists in the classification of research objects using numerous computational procedures. It makes it possible to classify objects by several features at the same time, to transform a large volume of versatile information into an ordered, compact form. The hierarchical clustering method and the k-means method are considered. The experiments carried out in SPSS Statistics made it possible to choose the k-means method as a method for further research. A computer experiment on the construction of a system of clusters was conducted using the data of the report of the Ministry of Digital Transformation of Ukraine – the final indicators (sub-indices) of the digital transformation of regions. Five clusters have been identified, which can be matched to the levels of development of digitalization of regions. A comparison of these results with the integral indices of the digital transformation of regions, which are determined by a different methodology, shows a good consistency. The method of k-means is planned to be chosen as the basis for the software implementation of the clustering subsystem of the decision support system for the analysis of the digital transformation of regions.

Key words: cluster analysis, digitization of regions, SPSS.

Актуальність.

Для вирішення проблем післявоєнного відновлення України значна увага приділяється цифровому розвитку економіки та суспільства. Цифровий розвиток регіонів має на меті забезпечити економічне зростання та покращення рівня життя громадян, зробити державні послуги більш доступними та якісними.

Для планів розбудови та визначення тенденцій потребується систематизація наявних знань щодо трендів та інструментів цифровізації регіональних систем.

До показників цифровізації можна віднести такі фактори, як рівень цифровізації економіки; охоплення домогосподарств цифровою сферою; інтенсивність державної участі у цифровізації та інші [1]. Для оцінки стану цифрових трансформацій регіонів потрібно залучати показники різного типу різних напрямів діяльності, як кількісні так і якісні, що формують базу багатовимірних даних для дослідження.

Україні провело дослідження індексу цифрової трансформації регіонів та представило звіт за результатами 2022 року [2]. Запропонований індекс являє собою один з інструментів вимірювання процесів цифровізації в регіонах України, дозволяє дослідити спроможність органів влади впроваджувати цифрові рішення та оцінювати рівень цифрової культури населення держави.

Для проведення аналізу, визначення трендів цифровізації та подальшого планування розвитку регіонів потрібно виділити спільні групи/кластери, щоб для них формувати відповідні пакети суспільно-економічних рекомендацій.

Також потрібно спиратись на характеристики цифровізації регіонів за попередній період, не тільки за результатами 2022 р. Нажаль, відсутній такий комплекс регіональних показників цифровізації раніше 2021 р. Доступні лише точкові статистичні дані Державної служби статистики України [3] щодо охоплення послугами інтернет, дані досліджень наявності ІТ-фахівців та деякі інші.

Тому є актуальним опис стану цифровізації регіонів України з їх ранжуванням, за наявними не повними даними попереднього періоду часу, та узгодження з результатами дослідження Міністерство цифрової трансформації 2022 р [2]. Це, можливо, також дозволить удосконалити методику визначення індексу цифрової трансформації регіонів.

Постановка завдання.

Метою даної роботи є дослідження та застосування кластерного аналізу для опису, обґрунтування трендів та показників узгодженості розвитку регіонів України в напрямі цифрових трансформацій.

Кластерний аналіз полягає у здійсненні класифікації об'єктів дослідження за допомогою численних обчислювальних процедур. Він дає можливість класифікувати об'єкти за декількома ознаками одночасно, перетворити великий обсяг різнобічної інформації в упорядкований, компактний вигляд. Це сприяє підвищенню рівня зрозумілості та сприйняття результатів аналізу, а також створює підґрунтя для прогнозування.

² The publication was prepared in the framework of the research project 0122U001232 "Restructuring of the national economy in the direction of digital transformations for sustainable development", funded by the National Research Foundation of Ukraine

Для досягнення цієї мети необхідно вирішити наступні завдання:

- Провести огляд літературних джерел щодо показників оцінки рівня цифрової трансформації та методів кластеризації.
- Вибрати можливі методи кластеризації та алгоритми досліджень
- Провести комп'ютерні експерименти з визначеними методами на наявних даних цифрового розвитку регіонів.
- Вибрати метод кластеризації для використання та програмної реалізації в підсистемі кластерного аналізу.
- Підготувати рекомендації для майбутніх досліджень з визначення трендів розвитку регіонів та взаємного зв'язку показників цифровізації та соціально-економічних показників розвитку регіонів України

Інформаційно-аналітичний огляд.

Індекс цифрової трансформації регіонів [2] є одним з інструментів вимірювання процесів інформатизації та цифровізації у регіонах. Загалом індекс містить 8 субіндексів (у рамках яких відображується 31 індикатор та 76 показників):

1. Інституційна спроможність
2. Розвиток інтернету
3. Розвиток ЦНАП
4. Режим «без паперів»
5. Цифрова освіта
6. Візитівка області
7. Проникнення базових послуг
8. Галузева цифрова трансформація

Більш детально про сутність субіндексів, методик визначення інтегрального індексу та результати розрахунків можна ознайомитись в матеріалах звіту [4].

Разом з тим, ця методика визначення індексу може бути удосконалена. Наприклад, шляхом урахування статистичних даних попередніх років щодо формування цифрового середовища: охоплення регіонів інтернет та засобами зв'язку, кількість ІТ-фахівців та ІТ-фірм, цифрова грамотність населення тощо

Аналіз основних тенденцій розвитку цифрової економіки в Україні за допомогою методів статистичного аналізу, синтезу, порівняння, систематизації та логічного узагальнення розглянуто в роботі [5]. У межах даної статті проаналізовано базові показники що характеризують готовність національної економіки до цифровізації (чисельність і структура підприємств, які мають доступ до мережі Інтернет, які мають вебсайт та інше) Забезпечення стабільного розвитку цифрової економіки вимагає якісного покриття Інтернетом, навчання висококваліфікованих фахівців по роботі з інформаційно-комунікаційними технологіями, розбудови новітньої цифрової інфраструктури та розвитку цифрових сервісів.

В науковій роботі [6] здійснено критичний аналіз результатів розрахунку індексу цифрової трансформації регіонів України за звітом Мінцифри, та зроблено висновок про фрагментарний підхід до формування вказаного інструменту вимірювання на рівні регіонів та про відсутність відповідної методики оцінки на рівні територіальних громад. Авторами розроблено свої пропозиції щодо удосконалення теоретико-методичного підґрунтя оцінки результативності й ефективності цифровізації регіонів і територіальних громад. Вони пропонують залучити пул індикаторів за шістьма напрямками цифровізації економіки у рамках бренду Дія, що характеризують тенденції розвитку даної сфери в Україні.

Метою роботи [7] є розробка концептуального базису оцінки й аналізу цифрового розвитку країни, який, на відміну від інших побудованих на базі методів інтелектуального аналізу багатовимірних об'єктів, що дозволяє підвищити якість формування та ухвалення управлінських рішень із діджиталізації найважливіших напрямків розвитку економіки України. У роботі розглянуто сутність цифровізації та виконано аналіз сучасного стану цифрової трансформації країни; проаналізовано особливості використання комплексу методів системного аналізу та математичного моделювання для оцінки стану цифровізації України. Результати може бути використано в практиці управління цифровим розвитком у державних та регіональних органах.

У статті [8] обґрунтовано доцільність проведення кластерного аналізу з метою дослідження відношення інноваційного процесу та економічного зростання. Оцінку рівня інноваційного розвитку соціально-економічних систем (регіонів та промислових інноваційних підприємств) здійснено на основі

методу багатовимірною статистичного аналізу (ієрархічного кластерного аналізу). У результаті проведеного дослідження за допомогою програмного забезпечення SPSS було виявлено кілька груп кластерів регіонів України.

Методи дослідження.

Кластерний аналіз є однією з основних методологій аналізу багатовимірних даних. Його використання широко поширене і швидко зростає [9]. Кластерний аналіз – це техніка для пошуку областей у n -вимірному просторі з великою концентрацією даних. Ці області називаються «кластерами». Зазвичай основним статистичним показником кластерного аналізу є центр цих кластерів [10].

Кластеризація (об'єднання в групи схожих об'єктів) є однією із фундаментальних задач Data Mining. Кластеризація включає в себе наступні етапи: виділення ознак, визначення метрики, розбиття об'єктів на групи та представлення результатів.

Кластерний аналіз не потребує апріорної інформації про дані та дозволяє розділити множину досліджуваних об'єктів на групи схожих об'єктів – кластери [11].

В роботі [9] задача кластеризації сформульована таким чином:

Є деяка кінцева множина об'єктів довільної природи $C = \{n_1, n_2, \dots, n_N\}$, кожний з яких $n_i = (x_{i1}, x_{i2}, \dots, x_{ik})$, $i = 1, \dots, N$, характеризується набором з K – ознак (вимірюваних характеристик об'єктів). Необхідно класифікувати ці об'єкти, тобто розбити їх множину на задану або довільну кількість груп (кластерів або класів) таким чином, щоб в кожену групу було включено об'єкти близькі між собою в тому чи іншому сенсі. Апріорна інформація про класифікацію об'єктів при цьому відсутня. Таким чином необхідно розбити множину векторів C на k попарно неперетинаючихся класів C_1, \dots, C_k так, щоб $\bigcup_{i=1}^k C_i = C$, де $1 < k < N$.

Для знаходження певного рішення даної задачі необхідно задати: міру схожості, спосіб кластеризації, встановити кількість кластерів.

За способами кластеризації методи кластерного аналізу можна розділити на дві великі групи: ієрархічні та неієрархічні методи [12]. Найбільш поширеними є **ієрархічні методи**, серед яких розрізняють агломеративний і дивізімний методи. На першому кроці **агломеративного методу** кожен об'єкт вважається окремим кластером. Два найбільш близьких об'єкта об'єднуються, і утворюється новий кластер. Процедура триває, доки всі об'єкти не будуть об'єднані в один кластер.

Для **дивізімного методу** спочатку всі об'єкти належать одному кластеру. Від цього кластера відокремлюються групи схожих між собою об'єктів. Так, на кожному кроці кількість кластерів зростає, а міра відстані між класами зменшується.

Другий підхід для процедури розбивки на кластери – ітеративні (неієрархічні) методи угруповання. Для застосування цього підходу необхідно є попередня розбивка даних на задане число кластерів і наступна робота з первинними даними. Ітеративна процедура починається з розбивки даних на задане число кластерів. Потім обчислюються центри ваги цих кластерів. Далі кожен об'єкт поміщають у той кластер, центр ваги якого є найближчим. Обчислюються нові центри ваги кластерів. Обчислення повторюються доти, поки кластери не стають стійкими, тобто перестануть змінюватися [13].

Найвідомішим алгоритмом неієрархічної кластеризації є сімейство алгоритмів K -means [14, 15]. Сутність цього методу полягає в наступному. Припустимо, що вже маємо гіпотези відносно кількості кластерів – тобто заздалегідь визначена кількість кластерів k , на які необхідно розбити наявні об'єкти. Серед множини об'єктів обирають k об'єктів в якості початкових центрів кластерів. Для кожного об'єкту розраховують відстані до центрів кластерів, і даний об'єкт відноситься до того кластера, відстань до якого виявилася мінімальною. Після чого для цього кластеру (в якому змінилася кількість об'єктів) розраховується нове положення центру кластеру (як середнє за кожною ознакою X_i) за всіма залученими до кластеру об'єктами. В загальному випадку, в результаті застосування методу k -середніх вихідна множина об'єктів розділюється рівно на k різних кластерів, розташованих на якнайбільших відстанях один від одного [9].

Відстань між об'єктами n_i та n_j може бути обчислена, наприклад, як Евклідова відстань

$$d(n_i, n_j) = \sqrt{\sum_{l=1}^K} \quad (1)$$

Для зберігання інформації про приналежність об'єкту до деякого кластеру в методі k -середніх вводиться матриця $U = \{u_{ij}\}$, $i = 1, \dots, N$ та $j = 1, 2$. В першому стовбці матриці U містяться індекси кластерів, до яких відносяться об'єкти даних в другому стовбці – відстані від об'єктів до відповідних центрів кластерів. Алгоритм k -середніх є ітераційним (рис. 1). Автоматичне групування об'єктів в кластери припиняється при виконанні одного з критеріїв зупинки автоматичного групування.

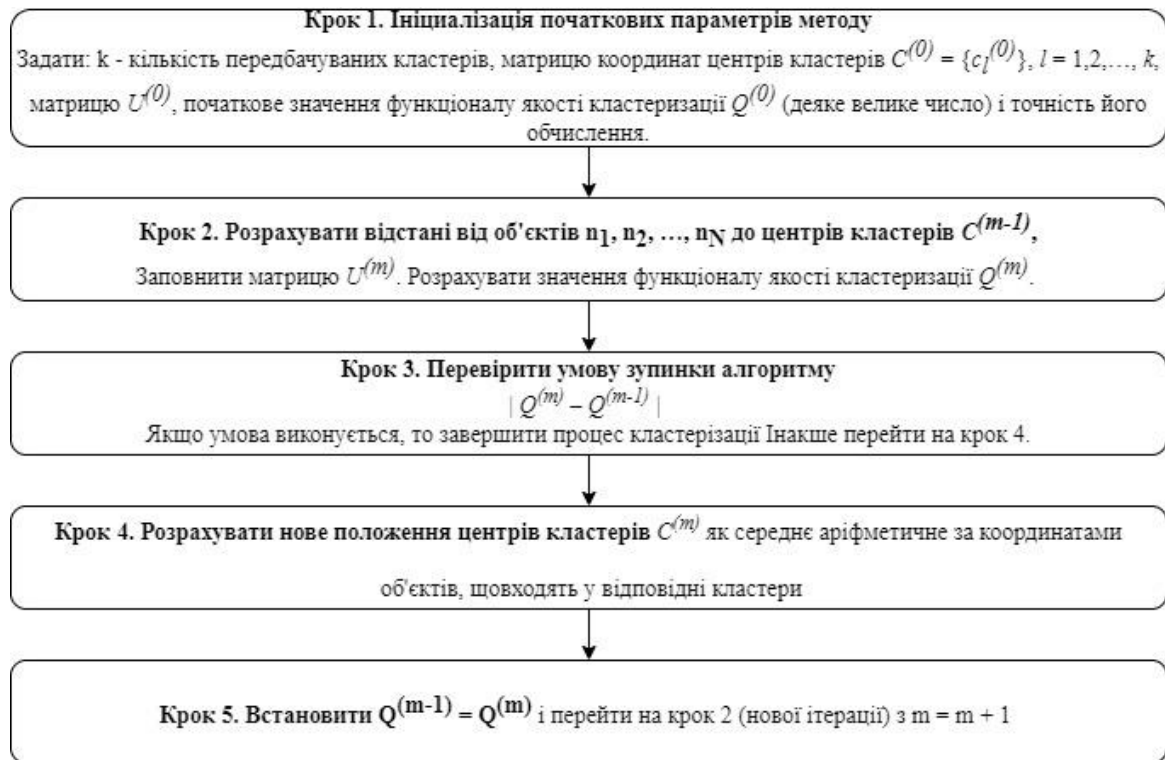


Рис. 1. Схема алгоритму кластеризації за методом k-середніх

Комп'ютерні експерименти по вибору метода дослідження.

В якості середовища проведення експериментів обрано SPSS Statistics – програмну платформу статистичного аналізу з потужним набором функцій [16]. SPSS Statistics дозволяє проводити кластеризацію ієрархічними та неієрархічними методами.

Вихідні дані для дослідження (таб.1) – підсумкові показники цифрової трансформації регіонів України 2022 року за даними Міністерства цифрової трансформації України [4].

Таблиця 1

Вихідні дані

| Область | Інституційна спроможність | Розвиток інтернету | Розвиток ЦНАП | Режим «без паперів» | Цифрова освіта | Візитівка області | Проникнення базових електронних послуг | Галузева цифрова трансформація |
|-------------------|---------------------------|--------------------|---------------|---------------------|----------------|-------------------|--|--------------------------------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Вінницька | 16 | 0,743 | 0,852 | 0,894 | 0,778 | 0,4 | 0,696 | 0,721 |
| Волинська | 0,6 | 0,733 | 0,911 | 0,928 | 0,256 | 0 | 0,772 | 0,716 |
| Дніпропетровська | 1 | 0,993 | 0,972 | 0,952 | 0,694 | 1 | 0,754 | 0,924 |
| Донецька | 0,325 | 0,252 | 0,631 | 0,527 | 0,396 | 0,68 | 0,596 | 0,345 |
| Житомирська | 0,38 | 0,909 | 0,895 | 0,859 | 0,349 | 0,4 | 0,646 | 0,403 |
| Закарпатська | 0,571 | 0,868 | 0,841 | 0,85 | 0,372 | 1 | 0,766 | 0,504 |
| Запорізька | 0,331 | 0,405 | 0,385 | 0,063 | 0,155 | 0,4 | 0,595 | 0,261 |
| Івано-Франківська | 0,81 | 0,896 | 0,853 | 0,162 | 0,124 | 0,8 | 0,78 | 0,338 |
| Київська | 0,71 | 0,377 | 0,824 | 0,711 | 0,258 | 1 | 0,61 | 0,21 |
| Кіровоградська | 0,213 | 0,388 | 0,664 | 0,531 | 0,163 | 0 | 0,614 | 0,154 |
| Луганська | 0,093 | 0,1 | 0,743 | 0,459 | 0,262 | 0,4 | 0,623 | 0,224 |
| Львівська | 0,89 | 0,833 | 0,822 | 0,924 | 0,562 | 0,8 | 0,68 | 0,82 |
| Миколаївська | 0,11 | 0,53 | 0,467 | 0,497 | 0,509 | 0,6 | 0,427 | 0,3 |

Закінчення табл. 1

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---------------|-------|-------|-------|-------|-------|------|-------|-------|
| Одеська | 0,9 | 0,841 | 0,705 | 0,958 | 0,702 | 1 | 0,877 | 0,808 |
| Полтавська | 0,8 | 0,894 | 0,905 | 0,967 | 0,88 | 0,6 | 0,737 | 0,56 |
| Рівненська | 1 | 0,907 | 0,875 | 0,891 | 0,195 | 0,8 | 0,705 | 0,573 |
| Сумська | 0,44 | 0,471 | 0,763 | 0,569 | 0,605 | 0 | 0,622 | 0,321 |
| Тернопільська | 1 | 0,992 | 0,83 | 0,933 | 0,718 | 1 | 0,834 | 0,999 |
| Харківська | 0,794 | 0,615 | 0,703 | 0,553 | 0,152 | 0,48 | 0,544 | 0,325 |
| Херсонська | 0,441 | 0,428 | 0,787 | 0,693 | 0,589 | 0,4 | 0,438 | 0,066 |
| Хмельницька | 0,861 | 0,609 | 0,719 | 0,653 | 0,165 | 0,4 | 0,596 | 0,458 |
| Черкаська | 0,744 | 0,771 | 0,855 | 0,803 | 0,312 | 0,6 | 0,759 | 0,387 |
| Чернівецька | 0,37 | 0,633 | 0,852 | 0,39 | 0,13 | 0,4 | 0,596 | 0,211 |
| Чернігівська | 0,493 | 0,612 | 0,629 | 0,596 | 0,267 | 0,4 | 0,4 | 0,507 |

Було проведено декілька експериментів по дослідженню методів кластеризації та кількості кластерів.

Ієрархічна кластеризація.

Ієрархічний метод використовує послідовне об'єднання об'єктів у кластери, малих кластерів у великі, що може бути візуально представлено у вигляді дерева вкладених кластерів – дендрограми (рис. 2). На графі дендрограми вздовж осі x є значення міри схожості, вздовж осі y – номери областей.

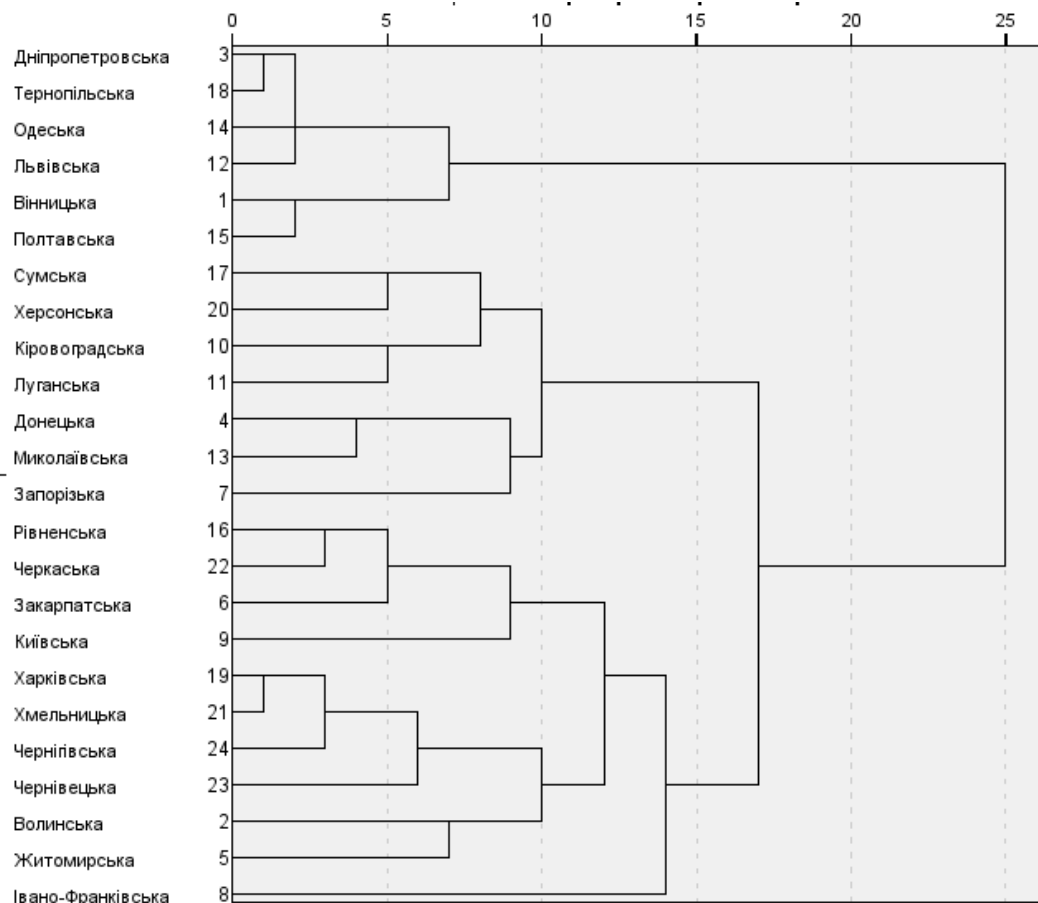


Рис. 2. Дендрограма поділу областей

Цей метод добре підходить на початковому етапі візуалізації та встановлення числа кластерів. На першому кроці в кластери об'єдналися (3, 18) та (19, 21). Потім утворився кластер (1, 15). Відстань між його елементами більше, ніж між елементами, які об'єдналися на першому кроці. Далі в один кластер об'єднуються (3, 18, 12). Аналогічним чином можна поєднати інші об'єкти в кластери. Процес завершується, коли всі об'єкти об'єднані в один кластер. Рішення по кількості кластерів приймається дослідни-

ком. Аналіз дендрограми (рис. 2) дозволяє виділити п'ять кластерів однорідних станів в спостережуваній сукупності даних: (3, 18, 14, 12, 1, 15), (17, 20, 10, 11), (4, 13, 7), (16, 22, 6, 9), (19, 21, 24, 23, 2, 5, 8).

Кластеризація к-середніми

Особливістю цього підходу є необхідність попереднього визначення числа кластерів. Експериментально було проведено кластеризацію для N=3, 4, 5. Результати наведено у таблиці 2.

Таблиця 2

Експериментальні виділення кластерів

| N=3 | | N=4 | | N=5 | |
|-------------------|---------|-------------------|---------|-------------------|---------|
| Область | Кластер | Область | Кластер | Область | Кластер |
| Донецька | 1 | Івано-Франківська | 1 | Івано-Франківська | 1 |
| Запорізька | 1 | Київська | 1 | Київська | 1 |
| Івано-Франківська | 1 | Харківська | 1 | Харківська | 1 |
| Київська | 1 | Хмельницька | 1 | Хмельницька | 1 |
| Кіровоградська | 1 | Черкаська | 1 | Черкаська | 1 |
| Луганська | 1 | Вінницька | 2 | Вінницька | 2 |
| Миколаївська | 1 | Волинська | 2 | Волинська | 2 |
| Харківська | 1 | Житомирська | 2 | Житомирська | 2 |
| Херсонська | 1 | Сумська | 2 | Дніпропетровська | 3 |
| Чернівецька | 1 | Дніпропетровська | 3 | Закарпатська | 3 |
| Чернігівська | 1 | Закарпатська | 3 | Львівська | 3 |
| Вінницька | 2 | Львівська | 3 | Одеська | 3 |
| Волинська | 2 | Одеська | 3 | Полтавська | 3 |
| Житомирська | 2 | Полтавська | 3 | Рівненська | 3 |
| Сумська | 2 | Рівненська | 3 | Тернопільська | 3 |
| Хмельницька | 2 | Тернопільська | 3 | Запорізька | 4 |
| Черкаська | 2 | Донецька | 4 | Кіровоградська | 4 |
| Дніпропетровська | 3 | Запорізька | 4 | Луганська | 4 |
| Закарпатська | 3 | Кіровоградська | 4 | Чернівецька | 4 |
| Львівська | 3 | Луганська | 4 | Донецька | 5 |
| Одеська | 3 | Миколаївська | 4 | Миколаївська | 5 |
| Полтавська | 3 | Херсонська | 4 | Сумська | 5 |
| Рівненська | 3 | Чернівецька | 4 | Херсонська | 5 |
| Тернопільська | 3 | Чернігівська | 4 | Чернігівська | 5 |

Для змістовного аналізу даних кластерів були побудовані діаграми (рис. 3-5).

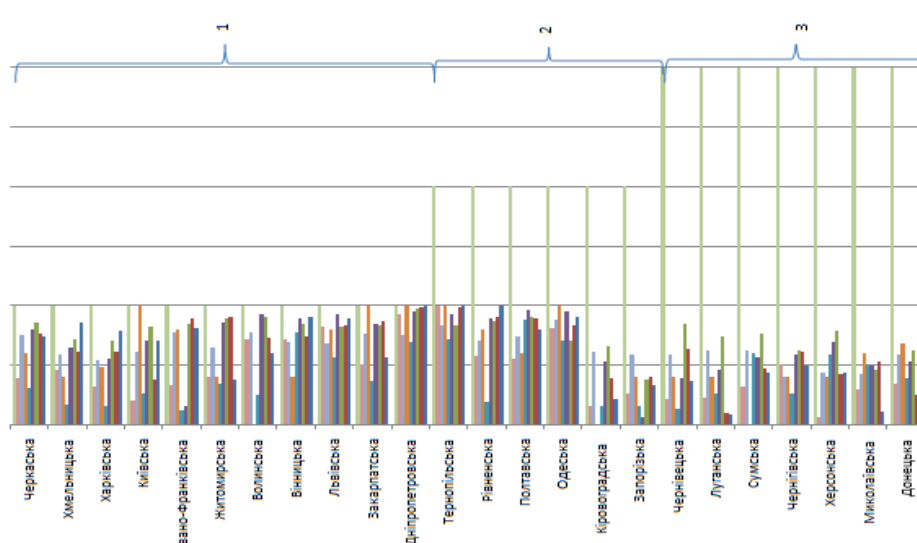


Рис. 3. Розбиття на кластери, N=3

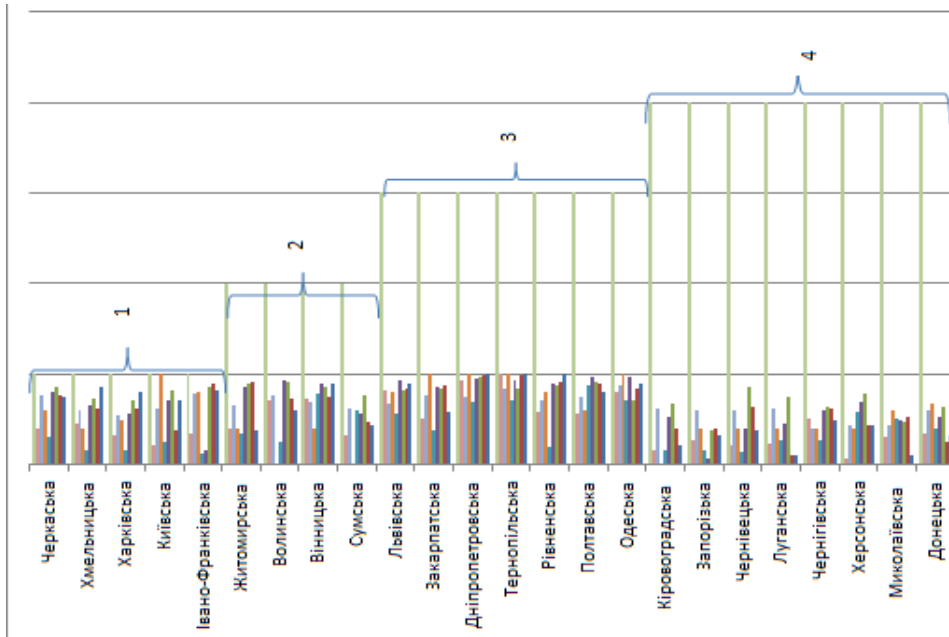


Рис. 4. Розбиття на кластери, N=4

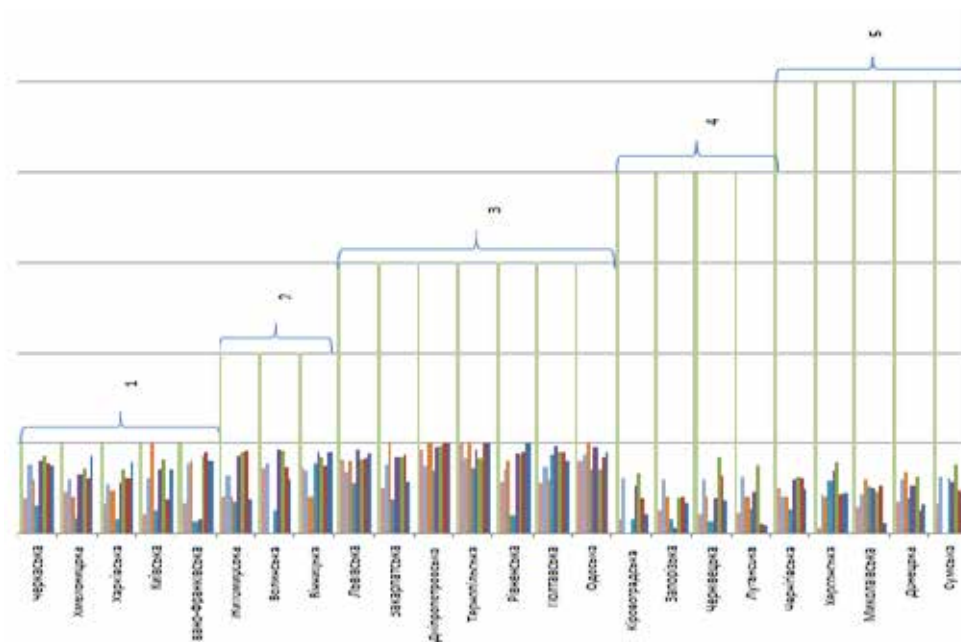


Рис. 5. Розбиття на кластери, N=5

На рис. 3 спостерігаємо доволі неоднорідне розподілення областей за показниками. Особливо в другому кластері. Збільшення числа кластерів (рис. 4) покращує загальну картину. Проте кластери 3-4 доволі неоднорідні. Варіант розбиття на п'ять кластерів дає найбільш логічно зрозумілу картину. Обрання цього варіанту дозволяє сформувати групи областей за такими рівнями показників: низький, нижче середнього, середній, вище середнього, високий.

Зведена таблиця 3 дає змогу оцінити кластери за рівнями показників. В ній наведені дані відхилення середнього значення по кластеру від середнього по Україні.

Таблиця 3

Зведена таблиця за середніми показниками для n=5

| кластер | Інституційна спроможність | Розвиток інтернету | Розвиток ЦНАП | Режим «без паперів» | Цифрова освіта | Візитівка області | Проникнення базових електронних послуг | Галузева цифрова трансформація | рівні |
|---------|---------------------------|--------------------|---------------|---------------------|----------------|-------------------|--|--------------------------------|------------------|
| 4 | 0,38 | 0,30 | 0,11 | 0,33 | 0,229 | 0,27 | 0,04 | 0,26 | низький |
| 5 | 0,27 | 0,22 | 0,11 | 0,11 | -0,06 | 0,15 | 0,15 | 0,16 | нижче середнього |
| 1 | -0,14 | 0,02 | -0,02 | 0,11 | 0,20 | -0,08 | 0,00 | 0,13 | середній |
| 2 | 0,01 | -0,11 | -0,11 | -0,20 | -0,05 | 0,30 | -0,05 | -0,13 | вище середнього |
| 3 | -0,24 | -0,22 | -0,07 | -0,23 | -0,18 | -0,31 | -0,11 | -0,26 | високий |

Ці дані візуалізовано на рис. 6.

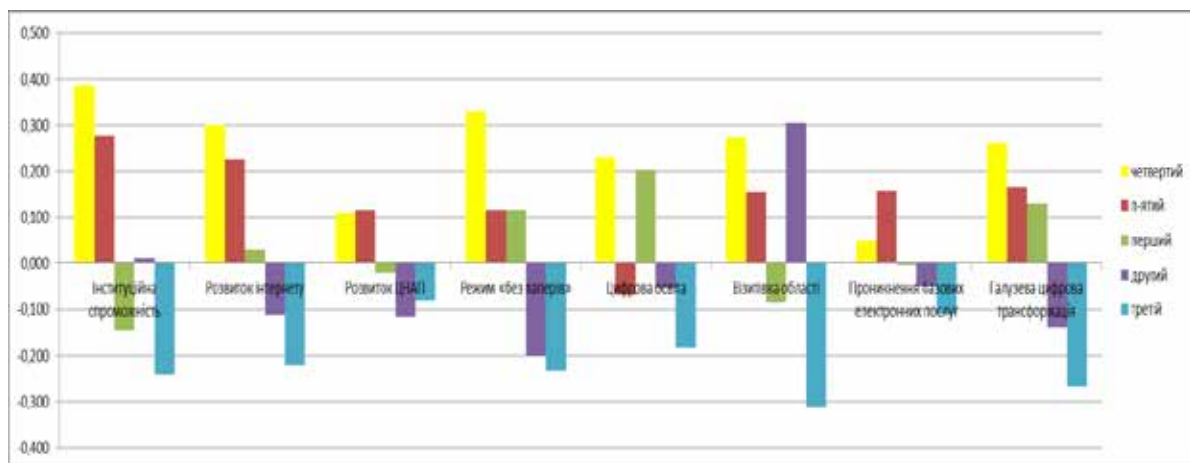


Рис. 6. Показники відхилення середніх значень по кластеру від середніх по Україні

На рисунку 7 жовтим кольором виділені мінімальні значення показників, фіолетовим – максимальні значення. Візуальний аналіз дозволяє виділити кластери, які містять найменші та найбільші значення за показниками. Так перший кластер (середній рівень) містить один найменший елемент (*Цифрова освіта*). Другий кластер (вище середнього) кращі показники – один найменший (*Візитівка області*) та один найбільший (*Розвиток ЦНАП*). Третій (високий рівень) містить всі інші найбільші значення. Кластер 5 (нижче середнього) містить два найменших значення (*Візитівка області* та *Галузева цифрова трансформація*). Четвертий кластер (низький рівень) містить решту найменших значень.

До **першого кластеру** надійшли області Івано-Франківська, Київська, Харківська, Хмельницька, Черкаська. В перший кластер увійшли регіони, які мають середні показники по Україні. Серед найгірших можна виділити *Цифрову освіту* та *Галузеву цифрову трансформацію*. Найкращим показником є *Інституційна спроможність* та *Розвиток ЦНАП*. Середні значення по *Розвитку Інтернету* та *Галузевій трансформації*. За 4 показниками області кластеру краще за середні показники по Україні. За 4 показниками області кластеру мають значення менше середнього по Україні.

До **другого кластеру** увійшли області Вінницька, Волинська, Житомирська. В другий кластер надійшли регіони, які мають рівень середній та вище середнього. Найкращий показник *Розвиток ЦНАП* та *Режим без паперів*. Найгірший *Візитівка області*. За 6 показниками області кластеру краще за середні показники по Україні. За 2 показниками області кластеру мають значення менше середнього по Україні.

До **третього кластеру** надійшли області з найкращими показниками. Дніпропетровська, Закарпатська, Львівська, Одеська, Полтавська, Рівненська, Тернопільська. Всі показники вище середніх.

| Область | Інституційна спроможність | | Розвиток інтернету | | Розвиток ЦНАП | | Режим «без паперів» | | Цифрова освіта | | Візитівка області | | Проникнення базових електронних послуг | | Галузева цифрова трансформація | | Класифікація | Показник Індексу |
|-------------------|---------------------------|-------|--------------------|-------|---------------|-------|---------------------|-------|----------------|-------|-------------------|------|--|-------|--------------------------------|-------|--------------|------------------|
| | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | | |
| Черкаська | | 0,744 | | 0,771 | | 0,855 | | 0,803 | | 0,312 | | 0,6 | | 0,759 | | 0,387 | 1 | 0,716 |
| Івано-Франківська | | 0,81 | | 0,896 | | 0,853 | | 0,162 | | 0,124 | | 0,8 | | 0,78 | | 0,338 | 1 | 0,683 |
| Хмельницька | | 0,861 | | 0,609 | | 0,719 | | 0,653 | | 0,165 | | 0,4 | | 0,596 | | 0,458 | 1 | 0,61 |
| Київська | | 0,71 | | 0,377 | | 0,824 | | 0,711 | | 0,258 | | 1 | | 0,61 | | 0,21 | 1 | 0,588 |
| Харківська | | 0,794 | | 0,615 | | 0,703 | | 0,553 | | 0,152 | | 0,48 | | 0,544 | | 0,325 | 1 | 0,571 |
| Вінницька | | 0,9 | | 0,743 | | 0,852 | | 0,894 | | 0,778 | | 0,4 | | 0,696 | | 0,721 | 2 | 0,769 |
| Волинська | | 0,6 | | 0,733 | | 0,911 | | 0,928 | | 0,256 | | 0 | | 0,772 | | 0,716 | 2 | 0,72 |
| Житомирська | | 0,38 | | 0,909 | | 0,895 | | 0,859 | | 0,349 | | 0,4 | | 0,646 | | 0,403 | 2 | 0,692 |
| Дніпропетровська | | 1 | | 0,993 | | 0,972 | | 0,952 | | 0,694 | | 1 | | 0,754 | | 0,924 | 3 | 0,916 |
| Тернопільська | | 1 | | 0,992 | | 0,83 | | 0,933 | | 0,718 | | 1 | | 0,834 | | 0,999 | 3 | 0,91 |
| Одеська | | 0,9 | | 0,841 | | 0,705 | | 0,958 | | 0,702 | | 1 | | 0,877 | | 0,808 | 3 | 0,836 |
| Полтавська | | 0,8 | | 0,894 | | 0,905 | | 0,967 | | 0,88 | | 0,6 | | 0,737 | | 0,56 | 3 | 0,814 |
| Львівська | | 0,89 | | 0,833 | | 0,822 | | 0,924 | | 0,562 | | 0,8 | | 0,68 | | 0,82 | 3 | 0,799 |
| Рівненська | | 1 | | 0,907 | | 0,875 | | 0,891 | | 0,195 | | 0,8 | | 0,705 | | 0,573 | 3 | 0,794 |
| Закарпатська | | 0,571 | | 0,868 | | 0,841 | | 0,85 | | 0,372 | | 1 | | 0,766 | | 0,504 | 3 | 0,756 |
| Чернівецька | | 0,37 | | 0,633 | | 0,852 | | 0,39 | | 0,13 | | 0,4 | | 0,596 | | 0,211 | 4 | 0,54 |
| Кіровоградська | | 0,213 | | 0,388 | | 0,664 | | 0,531 | | 0,163 | | 0 | | 0,614 | | 0,154 | 4 | 0,431 |
| Луганська | | 0,093 | | 0,1 | | 0,743 | | 0,459 | | 0,262 | | 0,4 | | 0,623 | | 0,224 | 4 | 0,404 |
| Запорізька | | 0,331 | | 0,405 | | 0,385 | | 0,063 | | 0,155 | | 0,4 | | 0,595 | | 0,261 | 4 | 0,37 |
| Сумська | | 0,44 | | 0,471 | | 0,763 | | 0,569 | | 0,605 | | 0 | | 0,622 | | 0,321 | 5 | 0,534 |
| Чернігівська | | 0,493 | | 0,612 | | 0,629 | | 0,596 | | 0,267 | | 0,4 | | 0,4 | | 0,507 | 5 | 0,522 |
| Херсонська | | 0,441 | | 0,428 | | 0,787 | | 0,693 | | 0,589 | | 0,4 | | 0,438 | | 0,066 | 5 | 0,5 |
| Донецька | | 0,325 | | 0,252 | | 0,631 | | 0,527 | | 0,396 | | 0,68 | | 0,596 | | 0,345 | 5 | 0,469 |
| Миколаївська | | 0,11 | | 0,53 | | 0,467 | | 0,497 | | 0,509 | | 0,6 | | 0,427 | | 0,3 | 5 | 0,431 |

Рис. 7. Порівняльна таблиця максимальних та мінімальних значень

До четвертого кластеру надійшли області з найгіршими показниками. Запорізька, Кіровоградська, Луганська, Чернівецька. Це кластер з найменшими показниками. Всі показники менші за середні.

В останній кластер надійшли області, показники в яких нижче середнього рівня. Донецька, Миколаївська, Сумська, Херсонська, Чернігівська. Тільки за показником *Цифрова освіта* даний кластер перевищує середні показники.

Обговорення результатів.

Проведено порівняння результатів наших досліджень з висновками звіту Індекс цифрової трансформації регіонів України [4]. За методикою Міністерства цифрової трансформації індекси регіонів визначались як сума середніх значень субіндексів з урахуванням вагових коефіцієнтів та вимірювались у відсотковому значенні. Результати порівняння дозволяють зробити висновок про майже точне співвідношення (табл. 4).

Таблиця 4

Порівняльна таблиця результатів кластеризації та показнику Індексу цифровізації

| Область | Кластер | Показник Індексу |
|-------------------|---------|------------------|
| 1 | 2 | 3 |
| Дніпропетровська | 3 | 0,916 |
| Тернопільська | 3 | 0,91 |
| Одеська | 3 | 0,836 |
| Полтавська | 3 | 0,814 |
| Львівська | 3 | 0,799 |
| Рівненська | 3 | 0,794 |
| Вінницька | 2 | 0,769 |
| Закарпатська | 3 | 0,756 |
| Волинська | 2 | 0,72 |
| Черкаська | 1 | 0,716 |
| Житомирська | 2 | 0,692 |
| Івано-Франківська | 1 | 0,683 |
| Хмельницька | 1 | 0,61 |
| Київська | 1 | 0,588 |
| Харківська | 1 | 0,571 |
| Чернівецька | 4 | 0,54 |
| Сумська | 5 | 0,534 |
| Чернігівська | 5 | 0,522 |

Закінчення табл. 4

| | | |
|----------------|---|-------|
| Херсонська | 5 | 0,5 |
| Донецька | 5 | 0,469 |
| Кіровоградська | 4 | 0,431 |
| Миколаївська | 5 | 0,431 |
| Луганська | 4 | 0,404 |
| Запорізька | 4 | 0,37 |

Результати кластеризації за к-середніх майже точно співвідносяться з показником Індексу. Сортування областей за рівнем показника Індексу цифровізації майже не змінило картину кластерів. Метод к-середніх дав логічно зрозумілі результати. Тому для програмної реалізації та подальших досліджень має сенс обрати саме цей метод.

Висновки та перспективи подальших досліджень.

Проведено інформаційно-аналітичний огляд досліджень стану цифровізації регіональних систем України, наявності методик визначення показників цифрових трансформацій. Найбільш комплексним та сучасним є дослідження Міністерства цифрової трансформації України з визначення індексу цифрової трансформації регіонів за результатами 2022 року. Хоча ця методика визначення індексу потребує удосконалення. Наприклад, потрібно урахування наявних статистичних даних попередніх років щодо формування цифрового середовища : охоплення регіонів інтернет та засобами зв'язку, кількість ІТ-фахівців та ІТ-фірм, цифрова грамотність населення тощо. Це дозволить покращити базу вхідних даних та виконувати аналіз трендів розвитку регіонів.

Для підготовки підґрунтя подальших досліджень узгодженості показників цифрових трансформацій регіонів з показниками суспільно-економічного розвитку, для визначення спільних груп/кластерів регіонів – щоб для них формувати відповідні пакети суспільно-економічних рекомендацій, був обраний кластерний аналіз.

На основі аналітичних міркувань та числових експериментів обрано метод кластеризації та алгоритм для програмної реалізації в підсистемі кластеризації системи підтримки прийняття рішень щодо аналізу цифрової трансформації регіонів.

Комп'ютерний експеримент з побудови системи кластерів проведено з використанням даних звіту Міністерства цифрової трансформації України – підсумкових показників (субіндексів) цифрової трансформації регіонів 2022 року. За результатами обчислень сформована система з п'яти кластерів. Порівняння цих результатів з інтегральними індексами цифрової трансформації регіонів [4], які визначені за іншою методикою, демонструє гарну узгодженість.

Це дозволяє рекомендувати запропоновану процедуру кластеризації для обробки наявних точкових статистичних даних попередніх років, виконання їх узгодження з результатами 2022 р. За результатом буде підготовлено фундамент, порівняльну базу для визначення трендів цифровізації та подальшого планування суспільно-економічного розвитку регіонів.

Список використаних джерел:

1. Цифрова економіка: тренди, ризики та соціальні детермінанти / за ред. О. Пішулиної. Київ : Центр Разумкова, «Заповіт», 2020. 274 с.
2. Результати цифрової трансформації в регіонах України. *Міністерство цифрової трансформації України*. URL: <https://thedigital.gov.ua/news/rezultati-tsifrovoi-transformatsii-v-regionakh-ukraini-1> (дата звернення: 01.08.2023).
3. Державна служба статистики України. URL: <https://ukrstat.gov.ua/> (дата звернення: 01.08.2023).
4. Індекс цифрової трансформації регіонів України. Підсумки 2022 року. *Міністерство цифрової трансформації України*. URL: <https://drive.google.com/drive/folders/1Wp61aHb0uRkb68mgebq8CvZbgVxupkCz> (дата звернення: 02.08.2023).
5. Котелевець Д.О. Тенденції розвитку цифрової економіки в Україні. *Проблеми сучасних трансформацій. Серія: економіка та управління*. 2022. № 5. DOI: 10.54929/2786-5738-2022-5-03-01 (дата звернення: 03.08.2023).
6. Рогозян Ю.С., Вахлакова В.В. Теоретико-методичні аспекти оцінки результативності й ефективності цифровізації економіки локальних територій України у воєнний і повоєнний час. *Академічні Візії*. 2023. № 19. URL: <http://dx.doi.org/10.5281/zenodo.7989723> (дата звернення: 04.08.2023).
7. Чаговець Л.О. Концептуальний базис оцінки й аналізу стану цифровізації України. *Digitalization and Information Society. Selected Issues. – Katowice: Publishing House of University of Technology*. 2023. С. 85–108. URL: <http://repository.hneu.edu.ua/handle/123456789/27670> (дата звернення: 03.08.2023).
8. Горященко Ю.Г. Застосування кластерного аналізу для оцінки інноваційного розвитку соціально-економічних систем. *Підприємництво і торгівля*. 2021. № 30. С. 25–32. DOI: 10.36477/2522-1256-2021-30-04.

9. Гавриленко О.В. Навчальний посібник з дисциплін Аналіз даних та Аналіз даних в управляючих системах : навчальний посібник. Київ: КПІ, 2020. – 85 с.
10. Zhao W., Chellappa R. *Face Processing: Advanced Modeling and Methods*. Elsevier, 2011. 768 p.
11. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining : навчальний посібник. Київ : КНУ ім. Т. Шевченка, 2017. 150 с.
12. Бізнес-аналітика багатовимірних процесів / Т.С. Клебанова та ін. : навчальний посібник. Харків: ХНЕУ ім. С. Кузнеця, 2018. 271 с.
13. Караєва Н.В., Варава І.А. Еколого-економічна оптимізація виробництва: методи та засоби кластерного аналізу : методичні вказівки. Київ : НТУУ «КПІ», 2016. 36 с.
14. Su, M.C.; Chou, C.H. A modified version of the K-means algorithm with a distance based on cluster symmetry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2001. Vol. 23, № 6. P. 674–680. DOI:10.1109/34.927466.
15. Ahmed M., Seraj R., Islam S. M.S. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics*. 2020. Vol. 9, № 8. P. 1–12. URL: <https://doi.org/10.3390/electronics9081295> (date of access: 03.08.2023).
16. Why IBM SPSS software? *IBM SPSS Software*. URL: <https://www.ibm.com/spss> (date of access: 03.08.2023).

References:

1. Pishhulyna, O. (Eds.). (2020). *Cyfrowa ekonomika: trendy, ryzyki ta socialni determinanty [Digital economy: trends, risks and social determinants]*. Kyiv: Centr Razumkova, «Zapovit» [in Ukrainian].
2. Sait Ministerstva cyfrowoi transformacii Ukrainy «Rezultaty cyfrowoi transformacii v regionah Ukrainy» [Site of the Ministry of Digital Transformation of Ukraine]. *thedigital.gov.ua*. Retrieved from <https://thedigital.gov.ua/news/rezultaty-sifrowoi-transformatsii-v-regionakh-ukraini-1> [in Ukrainian].
3. Sait Derzhavnoi sluzhby statystyky Ukrainy. [Sait of the State Statistics Service of Ukraine]. <https://ukrstat.gov.ua/>. Retrieved from <https://ukrstat.gov.ua/> [in Ukrainian].
4. Ministerstvo cyfrowoi transformacii Ukrainy «Indeks cyfrowoi transformacii regioniv Ukrainy. Pidsumky 2022 roku» [Ministry of Digital Transformation of Ukraine «Index of Digital Transformation of the Regions of Ukraine. Results of 2022»]. (n.d.). <https://drive.google.com/drive/folders/1Wp61aHb0uRKb68mgebq8CvZbgBxupkCz>. Retrieved from <https://drive.google.com/drive/folders/1Wp61aHb0uRKb68mgebq8CvZbgBxupkCz> [in Ukrainian].
5. Kotelevets, D.O. (2022). Tendencii rozvytku cyfrowoi ekonomiky v Ukrainy [Trends in the development of the digital economy in Ukraine]. *Problemy suchasnyh transformacij. Serija: ekonomika ta upravlinnja. – Problems of modern transformations. Series: Economics and Management*, 5. Retrieved from DOI: 10.54929/2786-5738-2022-5-03-01 [in Ukrainian].
6. Rogozjan Ju.S., & Vahlakova V.V. (2023). Teoretyko-metodychni aspekty ocinky rezultatyvnosti j efektyvnosti cyfrovizacii ekonomiky lokalnyh terytoryj Ukrainy u vojennyj i povojennyj chas [Theoretical and methodical aspects of assessing the effectiveness and efficiency of the economy digitalization for local territories of Ukraine in war and post-war period]. *Akademichni Vizii – Academy Vision*, 19. Retrieved from URL: <http://dx.doi.org/10.5281/zenodo.7989723> [in Ukrainian].
7. Chagovec L.O. (2022). Konceptualnyj bazys ocinky j analizu stanu cyfrovizacii Ukrainy [Conceptual basis of evaluation and analysis of the state of digitization of Ukraine]. *Digitalization and Information Society. Selected Issues. – Katowice: Publishing House of University of Technology*, 85-108. Retrieved from <http://repository.hneu.edu.ua/handle/123456789/27670> [in Ukrainian].
8. Horiashchenko, Yu.G. (2021). Zastosuvannja klasterneho analizu dlja ocinky innovacijnoho rozvytku socialno-ekonomichnyh system [Application of cluster analysis for assessment of innovative development of socio-economic systems]. *Pidpryjemnyctvo i torgivlja – Entrepreneurship and Trade*, 30, 25–32. Retrieved from DOI: 10.36477/2522-1256-2021-30-04 [in Ukrainian].
9. Gavrylenko, O.V. (2020). *Navchalnyj posibnyk z dyscyplin Analiz danyh ta Analiz danyh v upravljajuchyh systemah [Study guide for disciplines Data analysis and Data analysis in control systems]*. Kyiv: KPI [in Ukrainian].
10. Zhao, W., & Chellappa, R. (2011). *Face Processing: Advanced Modeling and Methods*. Elsevier.
11. Marchenko, O.O., & Rossada, T.V. (2017). *Aktualni problemy Data Mining [Actual problems of Data Mining]*. Kyiv: KNU T. Shevchenko [in Ukrainian].
12. Klebanova, T.S., Gurjanova, L.S., Chagovec, L.O., Panasenko, O.V., Sergijenko, O.A., & Jacenko, R.M. (2020). *Biznes-analzyka bagatovymirnyh procesiv [Business Analysis of Multidimensional Processes]*. Kharkiv: HNEU S. Kuznecja [in Ukrainian].
13. Karajeva, N.V., & Varava, I.A. (2016). *Ekologo-ekonomichna optymizacija vyrobnyctva: metody ta zasoby klasterneho analizu [Ecological and economic optimization of production: methods and resources of cluster analysis]*. Kyiv: NTUU «KPI» [in Ukrainian].
14. Su, M.C., & Chou, C.H. (2001). A modified version of the K-means algorithm with a distance based on cluster symmetry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. (Vol. 23), 6. 674–680. DOI: 10.1109/34.927466.
15. Ahmed, M., Seraj R., & Islam S. M.S. (2020). The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics*. (Vol. 9), 8. 1-12. URL: <https://doi.org/10.3390/electronics9081295>.
16. IBM SPSS Software. «Why IBM SPSS software?». *www.ibm.com*. Retrieved from URL: <https://www.ibm.com/spss>.