

# ОЦІНКА ЯКОСТІ КЛАСТЕРИЗАЦІЇ КООРДИНАТ ГЕОГРАФІЧНИХ ОБ'ЄКТІВ РІЗНИМИ МЕТОДАМИ

Дідик П.Ю., Максимова С.С.

Харківський національний університет радіоелектроніки

Україна, 61166, Харків, пр. Науки 14

E-mail: [pavlo.doidyk@nure.ua](mailto:pavlo.doidyk@nure.ua)

**Анотація:** У даній статті було розглянуто оцінку результати кластеризації населених пунктів різними методами, а саме K-means, DBSCAN, Agglomerative, Spectral, Berch. Проведено оцінку якості кожного методу кластеризації за допомогою відповідних критеріїв. Дано рекомендації щодо підбирання вхідних даних для кластеризації.

**Ключові слова:** кластеризація, метод, оцінка якості

## ASSESSMENT OF THE QUALITY OF CLUSTERING GEOGRAPHIC OBJECTS' COORDINATES USING VARIOUS METHODS.

Didyk P., Maksimova S.

Kharkiv National University of Radio Electronics

Ukraine, 61166, Kharkiv, Nauky Ave. 14

**Annotation:** In this article, an evaluation of clustering results for settlements using various methods, namely K-means, DBSCAN, Agglomerative, Spectral, and Berch, has been discussed. The quality assessment of each clustering method was conducted using respective criteria. Recommendations regarding the selection of input data for clustering have been provided.

**Key words:** clustering, method, quality assessment

Цілеспрямована оцінка якості кластеризації географічних об'єктів за допомогою різних методів є критично важливою для широкого кола дисциплін, включаючи геоінформатику, геодезію, маркетинг та інші сфери. Розуміння ефективності кластеризаційних методів у відношенні до географічних даних може сприяти поліпшенню аналізу просторової інформації та прийняттю стратегічних рішень у сферах, де важлива локалізація та геопросторові взаємозв'язки.

У контексті Індустрії 4.0, кластеризація географічних об'єктів відіграє ключову роль у вирішенні різних завдань, пов'язаних з оптимізацією процесів, аналізом даних та прийняттям рішень. Одним з основних аспектів є використання географічних даних для побудови ефективних стратегій розташування виробничих підприємств, складів та інфраструктури. Кластеризація географічних об'єктів дозволяє ідентифікувати регіони або зони з концентрацією підприємств та ресурсів, що сприяє ефективному управлінню логістичними та постачальними ланцюгами, а також забезпечує оптимальне розташування виробничих об'єктів для зниження витрат та збільшення продуктивності.

Крім того, кластеризація географічних об'єктів важлива для аналізу та прогнозування тенденцій розвитку ринків та регіонів. Шляхом групування різних географічних зон за подібними ознаками та параметрами, можна отримати інсайти щодо споживчих звичок, попиту на ринку, а також виявити можливі потенційні ризики або перспективи для бізнесу. Це дозволяє підприємствам бути більш адаптивними та гнучкими у відповіді на змінні умови ринку.

Кластеризація - це процес групування схожих об'єктів у підмножини, які називаються кластерами. Метою кластеризації є виявлення природних структур або патернів в даних шляхом групування об'єктів, які мають схожі характеристики або властивості.

Кластеризація географічних об'єктів може бути використана для вирішення практичних завдань, таких як планування маршрутів доставки, розміщення нових об'єктів, оптимізація логістичних та постачальних ланцюгів і багато іншого. Кластеризація надає можливість

систематизувати та аналізувати географічні дані для ефективного прийняття управлінських та стратегічних рішень.

Аналізуючи дані, отримані при кластеризації різних географічних об'єктів України за допомогою різних методів кластеризації та вивчаючи графіки з оцінками різних підходів, можна провести оцінку ефективності цих методів та зробити висновки для кожного з них.

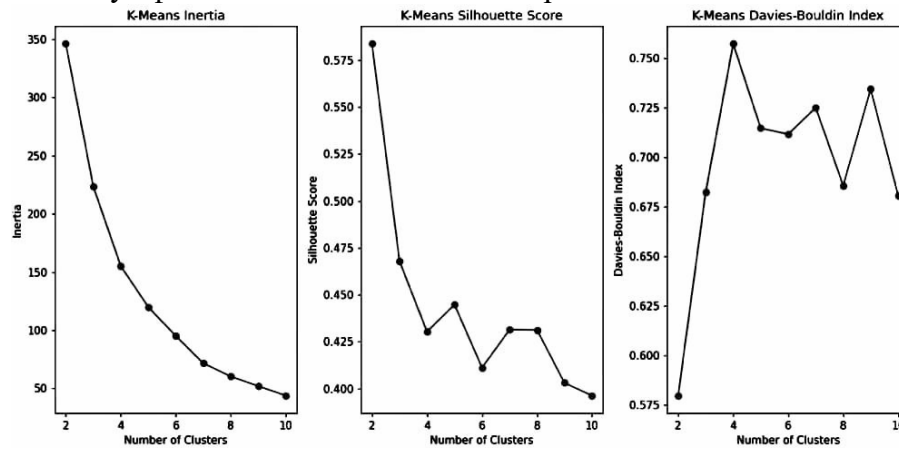


Рисунок 1 – Результат аналізу кластеризації методом K-means

а) Інерція свідчить про компактність точок всередині кластерів, тому оптимальним варіантом є вибір меншої інерції;

б) Вище значення коефіцієнта силуету вказує на більш точне відокремлення кластерів;

в) Найменше значення індексу Девіса-Боулдіна вказує на краще розділення кластерів;

Загалом, аналізуючи всі три методи оцінки, можна вважати, що оптимальною кількістю кластерів для даної задачі є 5.

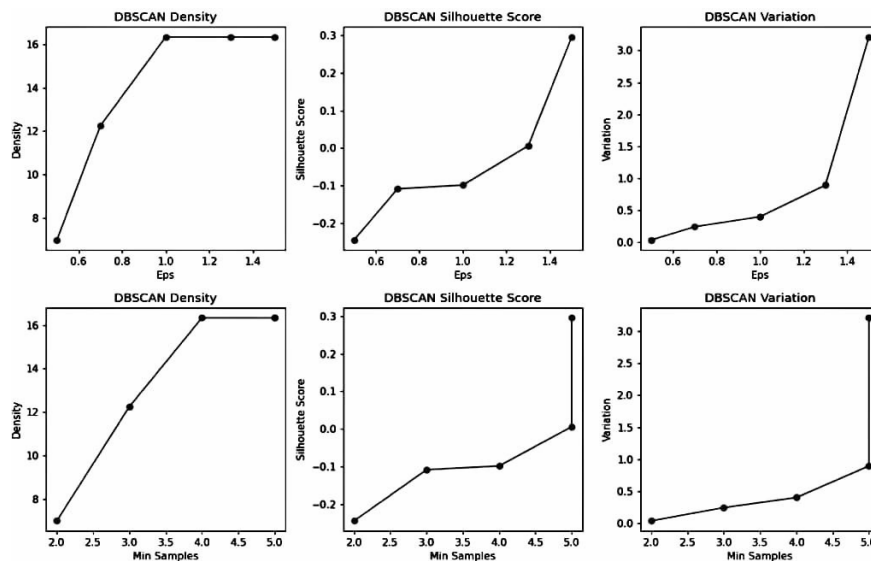


Рисунок 2 – Результат аналізу кластеризації методом DBSCAN

а) З поданих даних видно, що значення Density (eps) збільшується поступово, тоді як Density (min samples) залишається постійним. При eps=1.3 та min\_samples=4 спостерігається однакове значення для Silhouette Score та Variation, що робить ці параметри можливо оптимальними;

б) Негативні значення Silhouette Score, такі як -2.5 та -0.98, свідчать про низьку якість кластеризації.

в) Дані eps=1.5 та min\_samples=5 мають найвище значення Variation, що може свідчити про значну зміну внутрішньокластерних відстаней.

Загалом, при аналізі всіх трьох методів оцінки можна вважати, що оптимальними значеннями є  $\text{eps}=1.3$  та  $\text{min\_samples}=4$ .

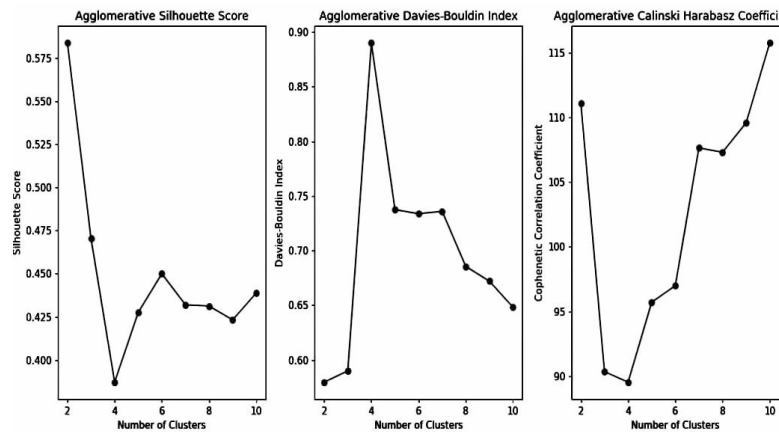


Рисунок 3 – Результат аналізу кластеризації методом Agglomerative

- а) Вищі значення Silhouette Score свідчать про вищу якість кластеризації;
- б) Нижчі значення Davies-Bouldin Index свідчать про кращу якість кластеризації;
- в) Вищі значення Calinski-Harabasz Index вказують на більш компактні та роздільні кластери;

Загалом, на основі цих показників розгляд може бути спрямований на 2 або 3 кластери як оптимальний вибір для цього набору даних..

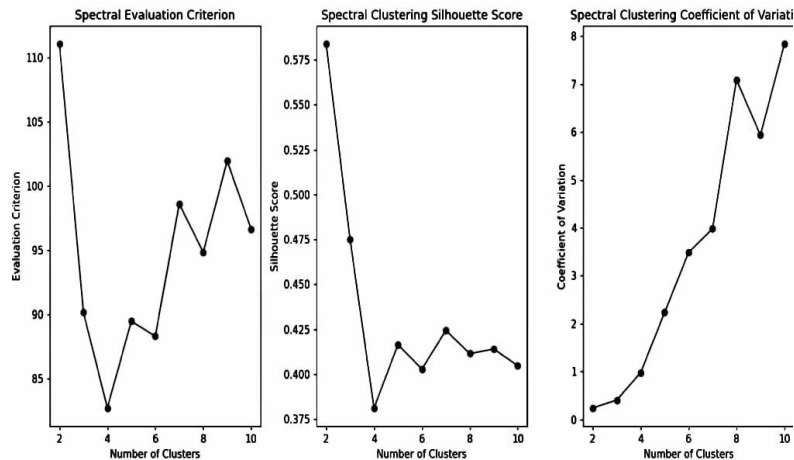


Рисунок 4 – Результат аналізу кластеризації методом Spectral

- а) за цими даними, ми можемо побачити, що значення критерію оцінки зменшується при збільшенні кількості кластерів від 2 до 4, а потім починає зростати;

- б) За цими даними, ми бачимо, що значення Silhouette максимальне при 2 кластерах і плавно знижується зі збільшенням кількості кластерів;

- в) за цими даними, ми бачимо, що значення коефіцієнта варіації збільшується зі збільшенням кількості кластерів;

Загалом, з урахуванням усіх трьох метрик, можна вважати оптимальним значенням кластерів 2 або 3. Рекомендується перевірити обидва варіанти та порівняти результати для визначення найкращого.

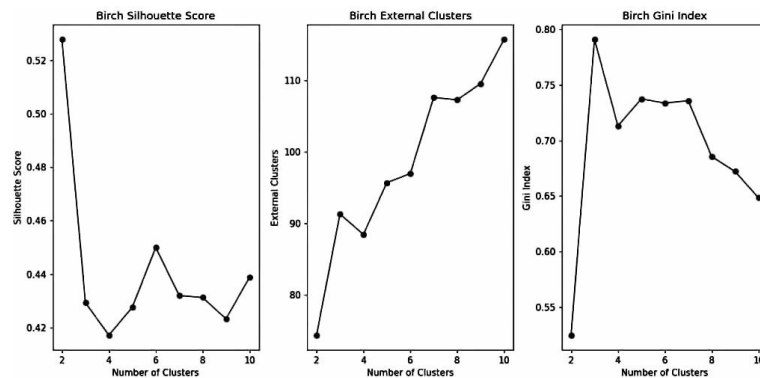


Рисунок 5 – Результат аналізу кластеризації методом Birch

а) Silhouette Score: ця метрика коливається від -1 до 1, де значення, що наближаються до 1, свідчать про вищу якість кластеризації;

б) External clusters: ця метрика вимірює зовнішню якість кластеризації, де більші значення вказують на кращу якість;

в) Gini Index: ця метрика коливається від 0 до 1, де значення, що наближаються до 0, вказують на вищу якість кластеризації;

Залежно від конкретних вимог або контексту задачі, ви можете обрати оптимальну кількість кластерів для методу Birch. За результатами оцінки, 10, 9 та 8 кластерів виявляються перспективними варіантами. Рекомендую спробувати ці значення і провести додатковий аналіз результатів для вибору найкращого варіанту.

**ВИСНОВКИ.** Аналіз показав, що кожен метод має свої переваги та обмеження, і вибір оптимального методу залежить від конкретних умов задачі. Результати дослідження підкреслили важливість глибокого розуміння кластеризаційних методів для вдосконалення аналізу географічних даних та розробки стратегій управління ресурсами та інфраструктурою у контексті Індустрії 4.0.

## ЛІТЕРАТУРА

1. Карпенко, А. А., Шевченко, В. І., & Григорович, А. П. (2019). Методи оцінки ефективності алгоритмів кластеризації в аналізі даних. *Кибернетика і системний аналіз*, 55(4), 161-176.

2. Лазарєв, О. В., & Попов, Є. П. (2020). Оцінка точності кластеризації текстових даних за допомогою внутрішніх критеріїв. *Інформаційні технології та комп'ютерна інженерія*, 4(46), 56-67.

3. Гончар, І. М., & Шуляк, В. В. (2018). Аналіз методів оцінки якості кластеризації в біологічних дослідженнях. *Біоінформатика та комп'ютерна біологія*, 9(1), 35-42.

4. Коваленко, Л. О., & Лисенко, Т. О. (2021). Використання валідаційних індексів для визначення кількості кластерів у задачі кластеризації даних. *Наукові записки НаУКМА*, 244, 58-65.

5. Літвінова, Т. В., & Іванов, О. О. (2019). Порівняльний аналіз метрик для оцінки якості кластеризації. *Вісник НТУУ "КПІ". Серія Інформатика, управління та обчислювальна техніка*, (78), 15-22.

6. Attar, H., & et al. (2022). Control System Development and Implementation of a CNC Laser Engraver for Environmental Use with Remote Imaging. *Computational Intelligence and Neuroscience*, 2022, Article ID 9140156, <https://doi.org/10.1155/2022/9140156>.

7. Attar, H., & et al. (2022). Zoomorphic Mobile Robot Development for Vertical Movement Based on the Geometrical Family Caterpillar. *Computational Intelligence and Neuroscience*, 2022, Article ID 3046116, <https://doi.org/10.1155/2022/3046116>.

8. Al-Sharo, Y., Abu-Jassar, A., Lyashenko, V., Yevsieiev, V., Maksymova, S. A Robo-hand prototype design gripping device within the framework of sustainable development, *Indian Journal of Engineering*, 20 2023 e37ije1673. <https://doi.org/10.54905/disssi.v20i54.e37ije1673>
9. Lyashenko, V., Abu-Jassar, A.T., Yevsieiev, V., Maksymova, S. Automated Monitoring and Visualization System in Production, *Int. Res. J. Multidiscip. Technovation*, 5(6) 2023 09-18. <https://doi.org/10.54392/irjmt2362>
10. Невлюдов І. Ш. ВЕАМ робототехніка : навч. посіб. / І. Ш. Невлюдов, В. В. Євсєєв, С. С. Максимова ; Харків. нац. ун-т радіоелектроніки, кафедра комп'ютерно-інтегрованих технологій, автоматизації та робототехніки (КІТАР). – Кривий Ріг : Видавець Чернявський Д. О., 2024. – 276 с. – ISBN 978-617-8045-79-1
11. Nevliudov, I., Yevsieiev, V., Maksymova, S., Klymenko, O. (2024), "The "load balancing" and "adaptive task completion" algorithms implementation on a pharmaceutical sorting conveyor line", *Innovative Technologies and Scientific Solutions for Industries*, No. 1 (24), P. 14–24. DOI: <https://doi.org/10.30837/ITSSI.2023.23.000>
12. Amer Abu-Jassar, Vladyslav Yevsieiev, & Svitlana Maksymova. (2024). The Optical Flow Method and Graham's Algorithm Implementation Features for Searching for the Object Contour in the Mobile Robot's Workspace. *Journal of Universal Science Research*, 2(3), 64–75.
13. Svitlana Maksymova, Vladyslav Yevsieiev, & Amer Abu-Jassar. (2024). Gripping Device Development: Some Aspects. *Journal of Universal Science Research*, 2(1), 150–158.
14. Svitlana Maksymova, & Vladyslav Yevsieiev. (2024). Coin Counting Device Kinematic Diagram Development. *Journal of Universal Science Research*, 2(1), 159–168.
15. Svitlana Maksymova, Vladyslav Yevsieiev, & Amer Abu-Jassar. (2024). The Bipedal Robot a Kinematic Diagram Development. *Journal of Universal Science Research*, 2(1), 6–17.
16. Svitlana Maksymova, Vladyslav Yevsieiev, & Amer Abu-Jassar. (2024). Gripping Device Development: Some Aspects. *Journal of Universal Science Research*, 2(1), 150–158.
17. Yevsieiev, V. Comparative Analysis of the Characteristics of Mobile Robots and Collaboration Robots Within INDUSTRY 5.0. / V. Yevsieiev, D. Gurin // In the VI International Scientific and Theoretical Conference, September 8, 2023. Chicago, USA. P.92-94
18. A Small-Scale Manipulation Robot a Laboratory Layout Development / Yevsieiev V., Starodubcev N., Maksymova S., Stetsenko K. // *International independent scientific journal*, №47, 2023. P.18-28.