

2. ТРАКТОРИ ТА АВТОМОБІЛІ. Електронний навчальний посібник: веб-сайт. URL: [https://elib.lntu.edu.ua/sites/default/files/elib\\_upload/%D0%A2%D1%80%D0%B0%D0%BA%D1%82%D0%BE%D1%80%D0%B8%20%D1%82%D0%B0%20%D0%B0%D0%B2%D1%82%D0%BE%20\(%D0%A1%D0%B0%D1%86%D1%8E%D0%BA%20%D0%92.%D0%92.\)/index.html](https://elib.lntu.edu.ua/sites/default/files/elib_upload/%D0%A2%D1%80%D0%B0%D0%BA%D1%82%D0%BE%D1%80%D0%B8%20%D1%82%D0%B0%20%D0%B0%D0%B2%D1%82%D0%BE%20(%D0%A1%D0%B0%D1%86%D1%8E%D0%BA%20%D0%92.%D0%92.)/index.html) (дата звернення: 12.04.2023)

3. Навісне обладнання для міні-трактора Види застосування: веб-сайт. URL: <https://gardenunion.com.ua/navisne-obladnannja-dlja-mini-traktora-vidi-zastosuvannja-36/> (дата звернення: 12.04.2023)

4. Харів Н. О. Бази даних та інформаційні системи: навчальний посібник. Рівне : НУВГП, 2018. – 127 с. URL: <https://ep3.nuwm.edu.ua/9129/3/%D0%A5%D0%B0%D1%80%D1%96%D0%B2%20%D0%9D.%D0%9E.pdf> (дата звернення: 19.04.2023)

*Грабовський Д. В., ст. гр. ММ-61-22*

*Ус М. О., ст. гр. ММ-41-20*

*Харківський національний автомобільно-дорожній університет*

## **ДОСЛІДЖЕННЯ ВПЛИВУ ПОХИБОК КЛАСТЕРИЗАЦІЇ ФРЕЙМІВ ДАНИХ ВИМІРЮВАНЬ НА ДОСТОВІРНІСТЬ ПРИЙНЯТТЯ РІШЕНЬ В ІНТЕЛЕКТУАЛЬНИХ ВИМІРЮВАЛЬНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ**

Похибки кластеризації даних вимірювань в інтелектуальних вимірювальних інформаційних системах виникають у процесі групування даних вимірювань, коли алгоритми кластеризації намагаються призначити дані вимірювань до різних кластерів. Як показали проведені дослідження основними причинами похибок кластеризації є [1-7]:

1. Неправильний вибір алгоритму. Різні алгоритми можуть давати різні результати для одних і тих самих даних. Вибір неправильного алгоритму може призвести до невірної кластеризації.

2. Перекривання кластерів. У деяких випадках кластери можуть бути сильно перекриваються, що робить важкою точну кластеризацію.

3. Наявність шуму. Шум в даних може призвести до неправильних кластерів, оскільки алгоритми можуть сприймати шумові точки як окремі кластери.

4. Недостатній або зайвий вибір ознак. Неправильний вибір ознак для кластеризації може спричинити похибки.

5. Параметризація алгоритмів. Налаштування параметрів алгоритмів (наприклад, кількість кластерів) може впливати на результати кластеризації.

6. Суб'єктивність. Інтерпретація та оцінка результатів кластеризації може бути суб'єктивною і призводити до різних варіантів.

Для зменшення похибок кластеризації важливо ретельно аналізувати дані, використовувати правильний алгоритм та належним чином налаштовувати параметри. Похибки кластеризації можуть впливати на достовірність прийняття рішень в інтелектуальних вимірjuвальних інформаційних системах. Якщо кластеризація ресурсів об'єктів у кластери виконується невірно, це може призвести до неправильних висновків та рішень на основі цих кластерів. Під час кластеризації частина інформації може бути втрачена, оскільки об'єкти групуються за певними ознаками, іноді ігноруючи інші важливі аспекти. Похибки в кластеризації можуть призвести до невпевненості у прийнятих рішеннях, оскільки важко визначити, наскільки кластери вірно відображають реальну структуру даних. Якщо приймаються рішення на основі невірної кластеризації, це може призвести до помилкових висновків та незадовільних результатів.

Тому важливо ретельно аналізувати результати кластеризації, розглядаючи їх у контексті конкретного завдання та забезпечуючи

адекватність та якість процесу кластеризації для забезпечення достовірності рішень.

Для зменшення похибок кластеризації важливо використовувати наступні методи та підходи [1-7]:

1. Вибір правильної кількості кластерів. Використання методів, таких як метод ліктя (Elbow method) або ієрархічна кластеризація, щоб підібрати оптимальну кількість кластерів.

2. Вибирати правильні ознаки для кластеризації. Важливо враховувати лише значущі ознаки для кластеризації та видаляти шумові або незначущі дані.

3. Нормалізація даних. Масштабування ознак може допомогти покращити якість кластеризації.

4. Використання відповідних алгоритмів для кластеризації. Вибір алгоритмів кластеризації, які найкраще підходять до вирішення поставленої задачі, таких як k-середніх (k-means), ієрархічна кластеризація, DBSCAN і т. д.

5. Валідація кластерів. Використовуйте метрики якості, такі як індекс силуету (Silhouette index) або інші, щоб оцінити якість кластеризації та вносити корективи в процес при необхідності.

6. Підбір ініціалізації. Правильний вибір початкового розташування центрів кластерів (у випадку k-середніх) може вплинути на кінцевий результат.

7. Розглядайте додаткові знання про дані вимірювань та специфіку завдання для покращення кластеризації.

8. Використовуйте ансамбльні методи, такі як ансамбль k-середніх (Ensemble k-means), для покращення стійкості та якості кластеризації.

За допомогою цих методів і підходів ви можете зменшити похибки кластеризації та отримати більш точні результати.

Література:

1. Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2017. “When should you adjust standard errors for clustering?” Working Paper. <http://www.nber.org/papers/w24003>.
2. Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. “Robust inference with multiway clustering.” *Journal of Business & Economic Statistics* 29 (2): 238–49. <http://www.jstor.org/stable/25800796>.
3. Donald, Stephen, and Kevin Lang. 2007. “Inference with difference-in-differences and other panel data.” *The Review of Economics and Statistics* 89 (2): 221–33. <https://doi.org/10.1162/rest.89.2.221>.
4. Erickson, Timothy, and Toni M. Whited. 2012. “Treating measurement error in Tobin’s q.” *Review of Financial Studies* 25 (4): 1286–1329. <https://doi.org/10.1093/rfs/hhr120>.
5. Sheppard, Kevin. 2023. “Linearmodels: Instrumental Variable and Linear Panel Models for Python.” <https://pypi.org/project/linearmodels/>.
6. Wasserstein, Ronald L., and Nicole A. Lazar. 2016. “The ASA Statement on p-Values: Context, process, and purpose.” *The American Statistician* 70 (2): 129–33. <https://doi.org/10.1080/00031305.2016.1154108>.
7. Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. The MIT Press. <http://www.jstor.org/stable/j.ctt5hhcfr>.

*Коваль Д. О., ст. гр. ММ-21-22*

*Харківський національний автомобільно-дорожній університет*

## **РИЗИКИ, ПОВ'ЯЗАНІ З ПРИЙНЯТТЯМ СТАТИСТИЧНИХ РІШЕНЬ ЗА РЕЗУЛЬТАТАМИ ВИМІРЮВАНЬ**

У зв'язку з ймовірнісним характером результатів вимірювань завжди існує невизначеність щодо вимірюваного значення параметра. Ця невизначеність відображається на правильності рішення, прийнятого на