

Міністерство освіти і науки України
Харківський національний автомобільно-дорожній університет

Механічний факультет
Кафедра метрології та БЖД

ДИПЛОМНА РОБОТА

магістра

ДОСЛІДЖЕННЯ ВПЛИВУ ПОХИБОК КЛАСТЕРИЗАЦІЇ
ФРЕЙМІВ ДАНИХ ВИМІРЮВАНЬ НА ДОСТОВІРНІСТЬ
ПРИЙНЯТТЯ РІШЕНЬ В БАГАТОПАРАМЕТРИЧНИХ
ПРОСТОРОВО РОЗПОДІЛЕНИХ ІНТЕЛЕКТУАЛЬНИХ
ВИМІРЮВАЛЬНИХ СИСТЕМАХ

Завідувач кафедри, канд. техн. наук, доц.



О. І. Богатов

Нормоконтролер, канд. техн. наук, доц



І. В. Грайворонська

Керівник, канд. техн. наук, доц



О. А. Коваль

Студент гр. ММ-61-22



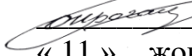
Д. В. Грабовський

Харків – 2023

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ АВТОМОБІЛЬНО-ДОРОЖНИЙ УНІВЕРСИТЕТ

Факультет механічний
Кафедра метрології та безпеки життєдіяльності
Освітній рівень другий (магістерський)
Спеціальність 152 Метрологія та інформаційно-вимірювальна техніка
Освітня програма «Інтелектуальні інформаційно-вимірювальні технології»

ЗАТВЕРДЖУЮ
Завідувач кафедри

 **О. І. Богатов**
« 11 » жовтня 2023 р.

ЗАВДАННЯ **НА ДИПЛОМНУ РОБОТУ СТУДЕНТУ**

Грабовському Данилу Владиславовичу

1. Тема роботи: "Дослідження впливу похибок кластеризації фреймів даних вимірювань на достовірність прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних системах".

Керівник роботи Коваль Олександр Андрійович, канд. техн. наук, доц.

Затверджені наказом Вченої ради механічного факультету від "11" жовтня 2023 р. № 130.



2. Строк подання студентом роботи 30.11.2023 р.

3. Вхідні дані до роботи: 1) Дані вимірювань трьох параметричної просторово розподіленої інтелектуальної вимірювальної інформаційної системи; 2) Результати оцінювання похибок кластеризації фреймів даних вимірювань в трьохпараметричній просторово розподіленій інтелектуальній вимірювальній інформаційній системі; 3) Критерії достовірності прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних системах.

4. Перелік питань, які потрібно розробити: 1. Вступ; 2. Методи кластеризації фреймів даних вимірювань в багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних системах; 3. Методика оцінювання похибок кластеризації фреймів даних вимірювань в багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних системах; 4 Дослідження впливу похибок кластеризації фреймів даних вимірювань на достовірність прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних системах; 5. Охорона праці і навколишнього середовища; 6. Висновок; 7. Перелік посилань.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень): плакати (слайди); 1 Результати досліджень у вигляді схем та графіків.

6. Консультанти по роботі, із зазначенням розділів проекту, що їх стосуються:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Завдання прийняв
5	Богатов О. І.		

7. Дата видачі завдання « 13 » вересня 2023 р.

КАЛЕНДАРНИЙ ПЛАН

Пор. №	Назва етапів дипломної роботи	Строк виконання етапів роботи	Примітка
1	Вивчення літератури	20.09.2023	виконано
2	Методи кластеризації фреймів даних вимірювань в багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних системах	30.09.2023	виконано
3	Аналіз похибок кластеризації фреймів даних вимірювань в багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних системах	01.10.2023	виконано
4	Дослідження впливу похибок кластеризації фреймів даних вимірювань на достовірність прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних системах	20.10.2023	виконано
5	Охорона праці і навколишнього середовища	01.11.2023	виконано
8	Формулювання висновків	10.11.2023	виконано
9	Оформлення пояснювальної записки	20.11.2023	виконано
10	Створення презентації у Power Point	27.11.2023	виконано
11	Подання роботи керівнику.	30.11.2023	виконано
12	Подання роботи на рецензію.	01.12.2023	виконано
13	Допуск до захисту.	05.12.2023	виконано

Студент



Грабовський Д. В.

Керівник роботи



Коваль О. А.

РЕФЕРАТ

Дипломна робота: 94 с., 29 рис., 1 додаток, 19 джерел.

ДОСТОВІРНІСТЬ, ІНТЕЛЕКТУАЛЬНА ВИМІРЮВАЛЬНА ІНФОРМАЦІЙНА СИСТЕМА, КЛАСТЕРИЗАЦІЯ, МАСИВИ ДАНИХ, ПОХИБКИ, ПРИЙНЯТТЯ РІШЕННЯ.

Об'єкт дослідження – просторово розподілена трьохканальна інтелектуальна вимірювальна інформаційна система тиску.

Мета роботи – дослідити вплив похибок кластеризації фреймів даних вимірювань на достовірність прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних системах.

Метод дослідження – кластерного, статистичного та часового аналізу даних вимірювань просторово розподіленої інтелектуальної вимірювальної інформаційної системи тиску.

Предмет дослідження – процес кластеризації вимірюваних масивів даних тиску.

В магістерській роботі проведені дослідження впливу похибок кластеризації фреймів даних вимірювань на достовірність прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних системах.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів.....	7
Вступ.....	8
1 Методи кластеризації фреймів даних вимірювань в багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних системах	12
1.1 Реалізація кластеризації k-середніх.....	13
1.2 Висновки до розділу.....	21
2 Аналіз похибок кластеризації фреймів даних вимірювань в багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних системах	23
2.1 Аналіз методу k-середніх.....	26
2.2 Аналіз методу t-SNE	28
2.3 Аналіз методу кластеризації на основі щільності ПКЩШ.....	30
2.4 Висновки до розділу.....	39
3 Дослідження впливу похибок кластеризації фреймів даних вимірювань на достовірність прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних системах	40
3.1 Особливості даних вимірювання тиску.....	43
3.1.1 Варіації метрики масивів вимірювання тиску.....	44
3.2 Метрики оцінювання похибок кластеризації.....	51
3.3 Однорідність, повнота і V-міра.....	53
3.4 Оцінювання впливу похибок кластеризації на достовірність прийняття рішення.....	56
3.5 Висновки до розділу.....	59
4 Охорона праці і навколишнього середовища.....	60
4.1 Виробнича санітарія.....	61
4.2 Безпека праці.....	66

4.3 Пожежна безпека.....	67
4.4 Розрахунок заземлюючого пристрою.....	70
4.5 Висновок до розділу.....	72
Висновки.....	73
Перелік посилань.....	75
Додаток А Ілюстративний матеріал до дипломної роботи	77

Кафедра МБЖД ХНАДУ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ,
ОДИНИЦЬ СКОРОЧЕНЬ І ТЕРМІНІВ

ARI – Adjusted Rand Index

PCA – Метод головних компонент

SOM – Апарат нейронних мереж

t-SNE – t-distributed stochastic neighbour embedding

АЕС – Атомна електростанція

ВДТ – Візуальні дисплейні термінали

ВЧ ПРА – Високочастотні пускорегулюючі апарати

ІВІС — Інтелектуальна вимірювальна інформаційна система

ПЕОМ – Персональна електронна обчислювальна машина

ПКЦШ – Просторова кластеризація даних на основі щільності з шумом

ПУЕ – Правила улаштування електроустановок

Кафедра МБЖД ХНАДУ

ВСТУП

Кластеризація даних вимірювань є важливим інструментом у сфері обробки даних і аналітики. Основна мета кластеризації полягає в групуванні подібних об'єктів чи спостережень у класи, або кластери, таким чином, щоб об'єкти в одному кластері були більш схожі між собою, ніж з об'єктами в інших кластерах. Ось деякі з головних причин, чому використовують кластеризацію даних вимірювань:

1. Групування та підсумовування даних. Кластеризація дозволяє легко визначити групи схожих об'єктів або вимірювань, що допомагає в створенні структурованого підсумкового представлення даних.

2. Візуалізація та розуміння. Кластеризація допомагає зменшити розмірність даних, що дозволяє легше візуалізувати і розуміти великі набори даних. Вона дозволяє аналітикам та дослідникам візуально виявляти залежності і закономірності в даних.

3. Класифікація та прогнозування. Кластеризація може слугувати підготовчим етапом для подальшого використання алгоритмів класифікації або прогнозування. Після групування схожих об'єктів можна використовувати моделі для передбачення класу або значення.

4. Виявлення відхилень і аномалій. Кластеризація дозволяє виявляти незвичайні або відхиляючі від стандартних паттернів групи даних. Це може бути корисним при виявленні аномалій або несподіваних вимірювань.

5. Спрощення задач аналізу. Групування схожих об'єктів спрощує завдання аналізу, оскільки можна працювати з представниками кожного кластера, а не з кожним окремим об'єктом.

6. Виділення суттєвих ознак. Кластеризація може виявити групи об'єктів з подібними характеристиками, що допомагає визначити суттєві ознаки та закономірності в даних.

Загалом, кластеризація даних вимірювань допомагає зрозуміти структуру даних, забезпечує основу для подальшого аналізу та допомагає приймати рішення на основі групового узагальнення даних.

Кластеризація даних вимірювань - це процес групування схожих об'єктів в одній категорії або кластері. Хоча кластеризація є потужним інструментом для аналізу даних, вона також може стикатися з рядом проблем. Деякі з основних проблем кластеризації даних вимірювань включають:

1. Визначення кількості кластерів (центроїдів). Обрання оптимальної кількості кластерів може бути непростим завданням. Визначення заздалегідь кількості кластерів часто є важким і може вимагати експертного досвіду.

2. Чутливість до початкового вибору центроїдів. Вибір початкових центроїдів може впливати на результат кластеризації. Неправильний вибір початкових точок може призвести до поганого розподілу кластерів.

3. Чутливість до форми та розмірів кластерів. Деякі алгоритми кластеризації можуть виявлятися менш ефективними в тих випадках, коли кластери мають неправильні форми або розміри.

4. Наявність шуму та викидів. Викиди або шум у даних можуть впливати на точність кластеризації, особливо у випадках, коли вони не враховані в алгоритмах.

5. Неоднорідність густини даних. Якщо щільність даних в різних областях нерівномірна, то деякі алгоритми можуть мати проблеми із визначенням коректних кластерів.

6. Залежність від обраного алгоритму. Різні алгоритми кластеризації можуть давати різні результати для одних і тих самих даних, і вибір правильного алгоритму може вимагати знань про властивості даних та завдання.

7. Проблеми з високоімовірнісними даними. У випадку великої кількості вимірів або високоімовірних даних може виникнути проблема "прокляття розмірності", коли алгоритми стають менш ефективними через зростання обчислювальної складності та розрідженості даних.

Для розв'язання цих проблем важливо вибрати алгоритми кластеризації відповідно до характеристик конкретних даних і враховувати їхні унікальні властивості. Також може бути корисним використовувати методи валідації та оцінки результатів кластеризації.

Похибки кластеризації фреймів даних вимірювань можуть впливати на достовірність прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних системах у різних аспектах. Ось кілька можливих впливів:

1. Неточність кластеризації. Якщо алгоритм кластеризації не ефективний або недостатньо точний, то об'єкти можуть бути неправильно призначені до кластерів. Це може призвести до втрати інформації про реальну структуру даних.

2. Вплив на вимірювання. Похибки у визначенні кластерів можуть впливати на самі вимірювання. Наприклад, якщо об'єкти з різних кластерів помилково призначаються до одного, це може порушити аналіз взаємодій між різними кластерами та призвести до неточних результатів.

3. Неспроможність виділення важливих підгруп. Якщо алгоритм кластеризації не може ефективно виділити важливі підгрупи даних, це може ускладнити виявлення складних залежностей і зробити неможливим або неточним аналіз параметрів, які дійсно впливають на систему.

4. Змішання різних джерел даних. У багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних системах можуть використовуватися дані з різних джерел. Похибки у кластеризації можуть призводити до неправильного змішування цих даних, що може впливати на якість прийняття рішень.

5. Неоднорідність у просторі. Якщо система має неоднорідний простір даних, де кластери мають різну щільність або форму, похибки кластеризації можуть призводити до неправильного розподілу важливих об'єктів.

6. Непрозорість рішень. Якщо похибки кластеризації не враховуються або не визначаються, це може зробити рішення системи менш прозорими і складними для розуміння та інтерпретації.

Уникнення та вирішення цих проблем включає в себе використання високоякісних алгоритмів кластеризації, уважний аналіз результатів, використання валідації та коригування похибок у процесі обробки даних в просторово розподілених інтелектуальних вимірювальних інформаційних системах.

Метою даної дипломної роботи є дослідження впливу похибок кластеризації фреймів даних вимірювань на достовірність прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних системах. Для досягнення цієї мети алгоритми кластеризації були реалізовані на Python з використанням хмарного сервісу Google Colab. В якості даних вимірювань були використані результати експериментальних досліджень трьохканальної просторово розподіленої інтелектуальної вимірювальної системи тиску, які були проведені Ковалем О. А. у 2021 році.

1 МЕТОДИ КЛАСТЕРИЗАЦІЇ ФРЕЙМІВ ДАНИХ ВИМІРЮВАНЬ В БАГАТОПАРАМЕТРИЧНИХ ПРОСТОРОВО РОЗПОДІЛЕНИХ ІНТЕЛЕКТУАЛЬНИХ ВИМІРЮВАЛЬНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

Для кластеризації фреймів даних вимірювань існує кілька методів. Основні з них:

1. К-середніх (K-Means).
2. Агломеративна ієрархічна кластеризація.
3. ПКЦШ (Просторова кластеризація даних на основі щільності з шумом).
4. Апарат нейронних мереж (SOM).
5. Метод головних компонент (PCA).

В даній дипломній роботі досліджувались методи кластеризації, які найбільш часто використовуються в інтелектуальних вимірювальних системах [1, 2]:

- К-середніх (K-Means).
- Агломеративна ієрархічна кластеризація.
- ПКЦШ (Просторова кластеризація даних на основі щільності з шумом).

К-середніх (K-Means) розділяє дані на К кластерів, призначаючи кожному елементу кластера згідно з його відстанню від центроїду [3].

Агломеративна ієрархічна кластеризація ґрунтується на припущенні, що кожен елемент є окремим кластером і послідовно об'єднує їх на основі визначеної метрики відстані.

ПКЦШ (Просторова кластеризація даних на основі щільності з шумом) це кластеризація на основі щільності, яка ідентифікує області великої щільності точок як кластери.

Апарат нейронних мереж (SOM) використовується для навчання без учителя і створює низькорозмірну репрезентацію вхідних даних.

Метод головних компонент (PCA) в основному використовується для зменшення розмірності. Він також може застосовуватися для кластеризації, виокремлюючи головні компоненти даних.

Метод слід вибирати відповідно до природи набору даних та цілей кластеризації.

1.1 Реалізація кластеризації k-середніх

Кластеризація – розбиття множини об'єктів на підмножини, звані кластерами. Кластеризація, будучи математичним алгоритмом, має широке застосування в багатьох сферах: починаючи з таких природничо-наукових галузей, як біологія і фізіологія, і закінчуючи маркетингом у соціальних мережах і пошуковою оптимізацією.

Існує безліч алгоритмів кластеризації, проте нижче буде розглянуто метод k-середніх, оскільки він є найбільш лаконічним і простим для розуміння.

При кластеризації K-середніх початковим завданням буде розподіл довільної кількості n-вимірних точок за k кластерами.

1. Випадковим чином створюються k точок, надалі називатимемо їх центрами кластерів;
2. Для кожної точки ставиться відповідно найближчий до неї центр кластера;
3. Обчислюються середні арифметичні точок, що належать до певного кластера. Саме ці значення стають новими центрами кластерів;
4. Кроки 2 і 3 повторюються доти, доки перерахунок центрів кластерів буде приносити плоди. Щойно вираховані центри кластерів збігатимуться з попередніми, алгоритм буде закінчено.

Приступимо до реалізації алгоритму. Вихідні дані алгоритму:

- n - кількість рядів;
- k - число кластерів;
- dim - розмірність точок (простору).

Результат роботи алгоритму алгоритму:

- `cluster` — двовимірний масив розмірності $\text{dim} * k$, що містить k точок
- центрів кластерів;
- `cluster_content` – масив, що містить k масивів, які є масивами точок, що належать відповідному кластеру.

Реалізуємо даний алгоритм на Python у вигляді функції:

```
def clusterization(array, k):
    n = len(array)
    dim = len(array[0])
    cluster = [[0 for i in range(dim)] for q in range(k)]
    cluster_content = [[] for i in range(k)]
    for i in range(dim):
        for q in range(k):
            cluster[q][i] = random.randint(0, max_cluster_value)
    cluster_content = data_distribution(array, cluster)
```

Змінні задано. Первинні центри кластерів створено за допомогою бібліотеки `random`, `max_cluster_value` - константа, що задає приблизні межі вихідної множини. За допомогою функції `data_distribution()` здійснено первинний розподіл точок за кластерами. Розглянемо цю функцію докладніше:

```
def data_distribution(array, cluster):
    cluster_content = [[] for i in range(k)]
    for i in range(n):
        min_distance = float('inf')
        suitable_cluster = -1
        for j in range(k):
            distance = 0
            for q in range(dim):
                distance += (array[i][q]-cluster[j][q])**2
            distance = distance**(1/2)
            if distance < min_distance:
                min_distance = distance
                suitable_cluster = j
        cluster_content[suitable_cluster].append(array[i])
    return cluster_content
```

В кожному рядку вираховується відстань до кожного центру кластерів. Тут застосовується стандартний алгоритм:

1. За початкову найкоротшу відстань (`min_distance`) береться несумірно велике зі значеннями точок число;

2. Потім проводиться обчислення відстані до центру кожного кластера. Обчислення відстані між точкою і центром у n -вимірному просторі проводиться за формулою:

$$p = \sqrt{\sum_{i=1}^n (x_n - x_{n \text{ centr}})^2}, \quad (1.1)$$

де p – відстань до центру кожного кластера;

x_n – n -й результат вимірювання тиску у вхідному масиві даних;

$x_{n \text{ centr}}$ – відстань між x_n і центром кластера.

3. Якщо обчислена відстань менша за мінімальну, то мінімальна відстань порівнюється до обчисленої і точка прив'язується до цього кластера (`suitable_cluster`);

4. Після опрацювання точки, в масив `cluster_content` в обраний кластер (`suitable_cluster`) кластер вкладається значення точки.

Функція повертає масив `cluster_content`. Надалі, як і годиться, ця послідовність дій повторюється:

```
previous_cluster = copy.deepcopy(cluster)
while 1:
    cluster = cluster_update(cluster, cluster_content, dim)
    cluster_content = data_distribution(array, cluster)
    if cluster == previous_cluster:
        break
    previous_cluster = copy.deepcopy(cluster)
```

Цей цикл цілісним чином описує крок 4 з опису алгоритму k-середніх. Після розподілу точок за центрами кластерів відбувається перерозподіл уже центрів кластерів за прив'язаними до них точками. Розглянемо функцію `cluster_content()` докладніше:

```
def cluster_update(cluster, cluster_content, dim):
    k = len(cluster)
    for i in range(k): #по i кластерам
        for q in range(dim): #по q параметрам
            updated_parameter = 0
            for j in range(len(cluster_content[i])):
                updated_parameter += cluster_content[i][j][q]
            if len(cluster_content[i]) != 0:
                updated_parameter = updated_parameter /
len(cluster_content[i])
            cluster[i][q] = updated_parameter
    return cluster
```

В кожному кластері, для кожного з n вимірів обчислюється нове значення за допомогою нехитрого середнього арифметичного: складаються всі значення і сума ділиться на кількість точок у кластері; кластер приймає оновлене значення. На цьому місці алгоритм закінчує свою роботу. Повний алгоритм має такий вигляд:

```
def clusterization(array, k):
    n = len(array)
    dim = len(array[0])
    cluster = [[0 for i in range(dim)] for q in range(k)]
    cluster_content = [[] for i in range(k)]
    for i in range(dim):
        for q in range(k):
            cluster[q][i] = random.randint(0, max_cluster_value)
    cluster_content = data_distribution(array, cluster)
    previous_cluster = copy.deepcopy(cluster)
    while 1:
        cluster = cluster_update(cluster, cluster_content, dim)
        cluster_content = data_distribution(array, cluster)
        if cluster == previous_cluster:
            break
    previous_cluster = copy.deepcopy(cluster)
```


Візуалізація кластерів. Візуалізуємо результат алгоритму для 3-х і 2-х мірного вихідних просторів:

```
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d
import numpy as np
```

Візуалізація для 2-х мірного простору відбувається таким чином:

```
def visualisation_2d(cluster_content):
    k = len(cluster_content)
    plt.grid()
    plt.xlabel("x")
    plt.ylabel("y")
    for i in range(k):
        x_coordinates = []
        y_coordinates = []
        for q in range(len(cluster_content[i])):
            x_coordinates.append(cluster_content[i][q][0])
            y_coordinates.append(cluster_content[i][q][1])
        plt.scatter(x_coordinates, y_coordinates)
    plt.show()
```

Створимо сітку, підпишемо назви осей. Потім у масиви, що відповідають осям, вкладаються значення точок. Після такої операції для кожного кластера викликається функція `scatter()` – розкид точок по площині. Наприкінці викликається функція відображення – `show()`.

Отримані результати кластеризації графічно відображені на рис.1.1, 1.2 та рис.1.3.

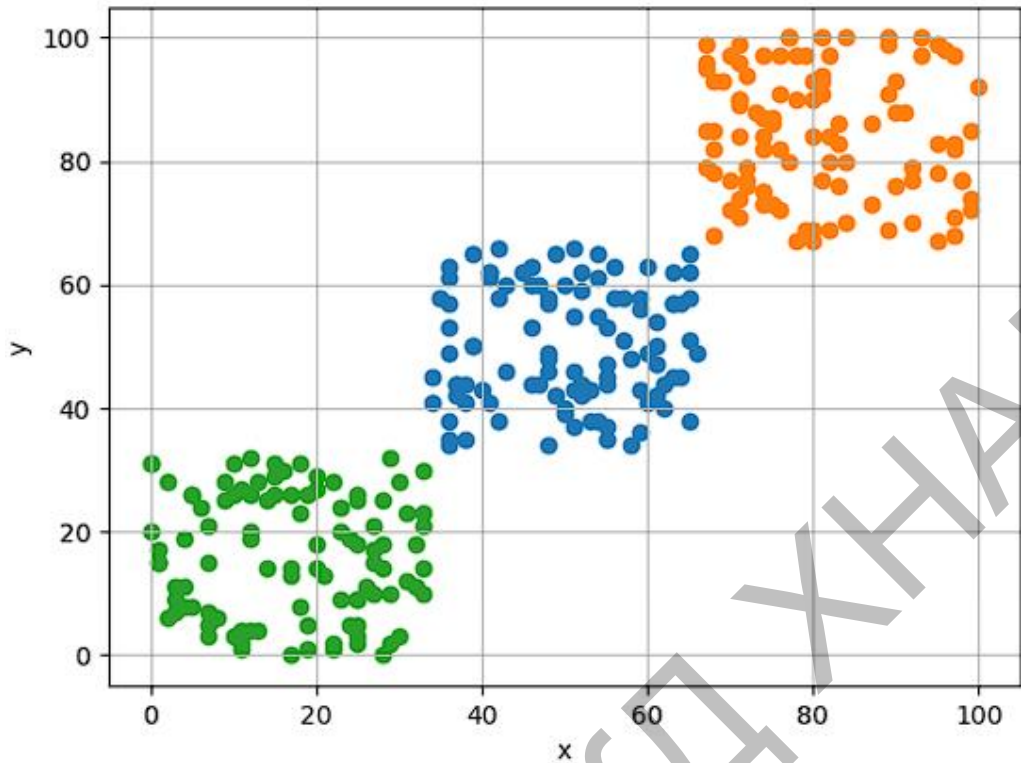


Рисунок 1.1 – Кластеризація для точок двовимірного простору.
Кількість точок – 300. Кількість кластерів – 3. Дані згруповано.

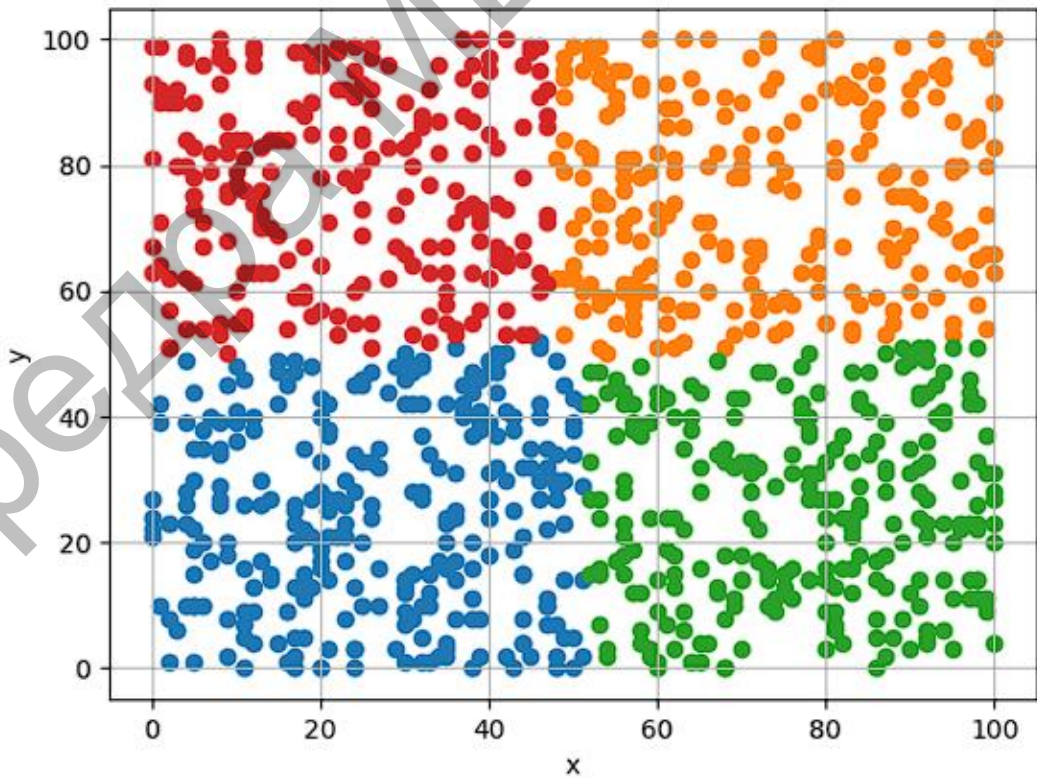


Рисунок 1.2 – Кластеризація точок двовимірного простору.
Кількість точок – 1000. Кількість кластерів – 4. Дані не згруповані.

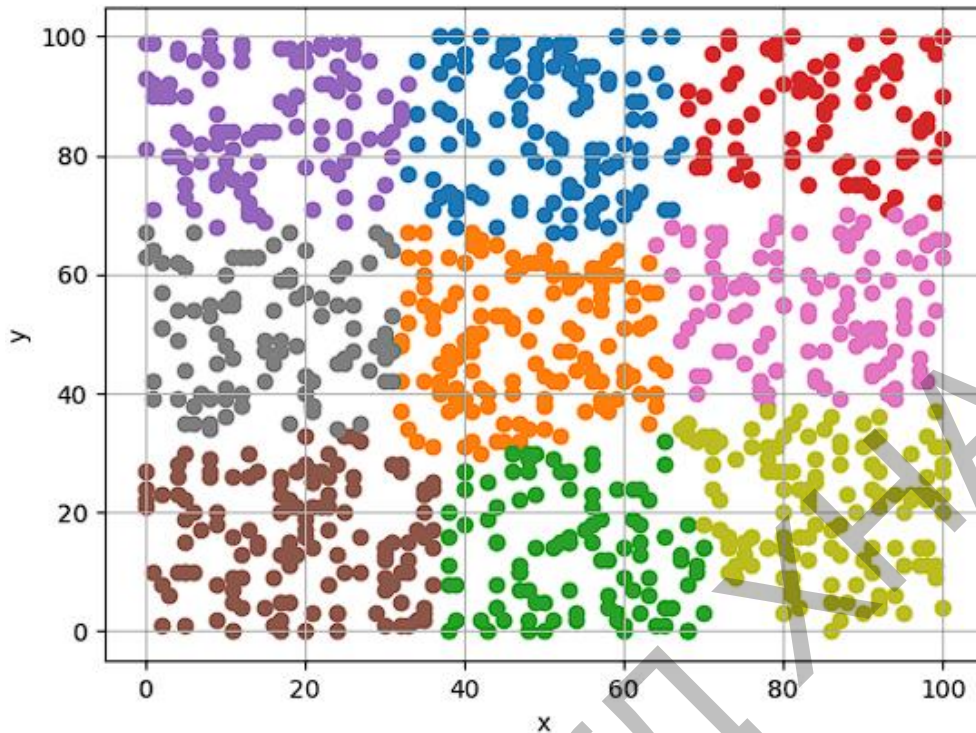


Рисунок 1.3 – Кластеризація точок двовимірного простору.

Кількість точок 1000. Кількість кластерів – 9. Дані не згруповані.

Аналогічним чином візуалізується результат алгоритму для 3-х мірного простору (рис. 1.4 – 1.6):

```
def visualisation_3d(cluster_content):
    ax = plt.axes(projection="3d")
    plt.xlabel("x")
    plt.ylabel("y")
    k = len(cluster_content)
    for i in range(k):
        x_coordinates = []
        y_coordinates = []
        z_coordinates = []
        for q in range(len(cluster_content[i])):
            x_coordinates.append(cluster_content[i][q][0])
            y_coordinates.append(cluster_content[i][q][1])
            z_coordinates.append(cluster_content[i][q][2])
        ax.scatter(x_coordinates, y_coordinates, z_coordinates)
    plt.show()
```

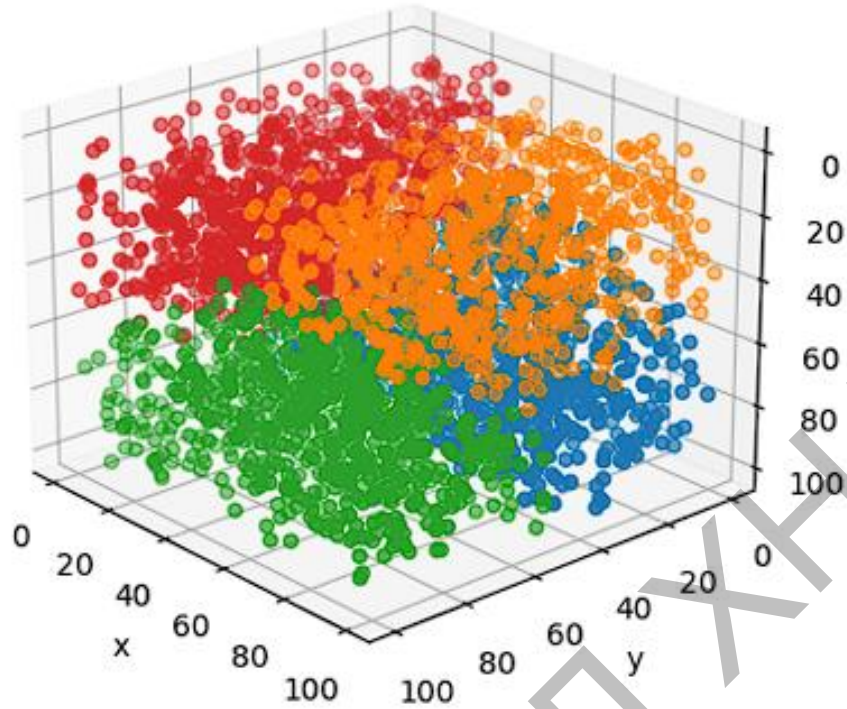



Рисунок 1.4 – Кластеризації для 3-х мірного простору.

Кількість точок – 3000. Кількість кластерів – 4. Дані не згруповані.

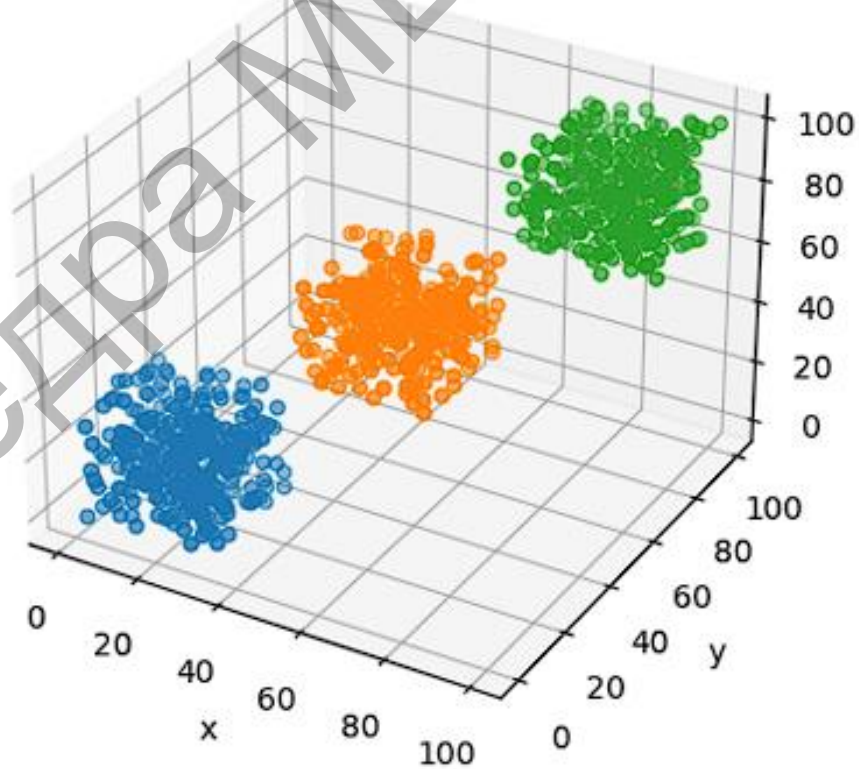


Рисунок 1.5 – Кластеризації тривимірного простору для згрупованих даних.

Кількість точок – 900. Кількість кластерів – 3

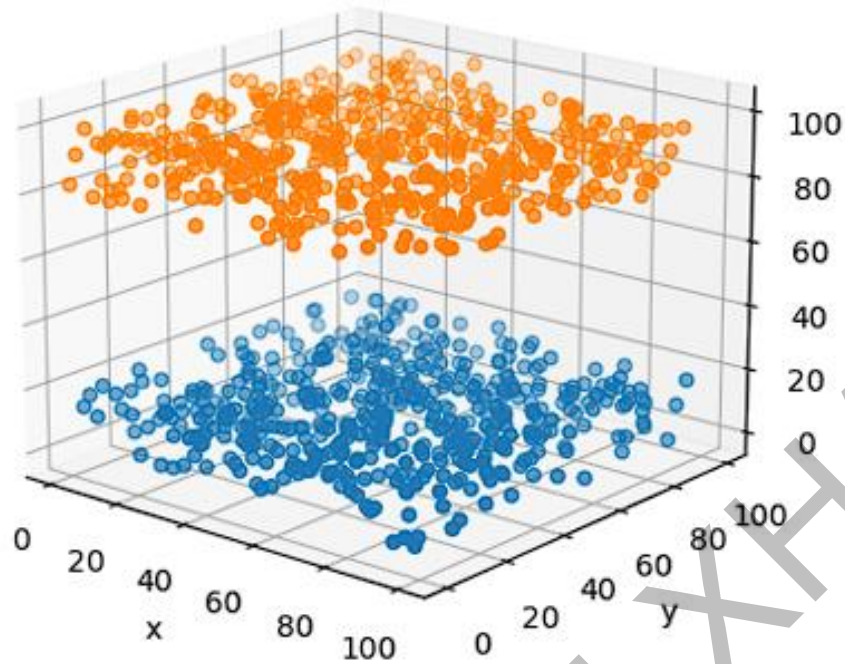


Рисунок 1.6 – Кластеризація тривимірного простору.

Кількість точок – 1000. Кількість кластерів – 2. Дані згруповано.

Таким чином, ми виконали необхідні й достатні умови для аналізу результатів кластеризації.

1.2 Висновки до розділу

Для успішного аналізу результатів кластеризації важливо мати необхідні й достатні умови. Ось деякі ключові аспекти, які слід враховувати:

1. Метрики якості кластеризації. Необхідно використовувати відповідні метрики для оцінки якості кластеризації.

2. Візуалізація результатів. Необхідно візуалізувати дані та кластери для отримання загального уявлення про те, наскільки добре вони розділені.

3. Параметри кластеризації. Необхідно перевіряти, чи правильно вибрані параметри для алгоритму кластеризації (наприклад, кількість кластерів). Використання неправильних параметрів може призвести до неправильних результатів.

4. Стабільність результатів. Необхідно перевіряти стабільність результатів кластеризації, використовуючи різні підвибірki або запуски алгоритму з різними початковими умовами. Стабільність може підтвердити адекватність отриманих кластерів.

5. Застосування отриманих кластерів. Завжди потрібно перевіряти, що отримані кластери можуть бути використані для вирішення задачі класифікації даних вимірювань у просторово розподілених вимірювальних інформаційних системах.

Загалом, тільки комбінація кількох аспектів дозволить здійснити комплексний аналіз результатів кластеризації і визначити їхню ефективність.

Кафедра МБЖД

2 АНАЛІЗ ПОХИБОК КЛАСТЕРИЗАЦІЇ ФРЕЙМІВ ДАНИХ ВИМІРЮВАНЬ В БАГАТОПАРАМЕТРИЧНИХ ПРОСТОРОВО РОЗПОДІЛЕНИХ ІНТЕЛЕКТУАЛЬНИХ ВИМІРЮВАЛЬНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

При аналізі похибок кластеризації фреймів даних вимірювань в багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних систем будемо використовувати результати експериментальних досліджень трьохканальної просторово розподіленої інтелектуальної вимірювальної системи тиску, які були проведені Ковалем О. А. у 2021 році. Основу вимірювального каналу тиску становили датчики тиску Сапфір-22ДТ-Вн (2440) (рис. 2.1). Вони мають наступні характеристики. Перетворювач вимірювальний різниці тисків Сапфір-22ДТ-Вн, модель 2440 з верхньою межею вимірювань 250 кПа і гранично допустимим робочим надлишковим тиском 16 МПа. Мембранний, з мембранами зі сплаву 36НХТЮ, з фланцями з вуглецевої сталі, що має вид кліматичного виконання УХЛЗ.1, робоча температура – від 5 °С до 50 °С з межею допустимої основної похибки $\gamma = \pm 0,25 \%$, з вихідним сигналом від 0 мА до 5 мА, з ніпелями, варіант встановлення перетворювача – на плиті, з вентильним блоком.



Рисунок 2.1 – Аналоговий датчик тиску Сапфір-22ДТ-Вн (2440)

Для складання прогнозів будемо використовувати дані експериментів. Датасет являє собою набір зі 150 записів із п'ятьма атрибутами в такому порядку: середнє значення вимірної величини (mean), дисперсія (variation), рівень шумів (noise level), відношення сигнал/шум (signal to noise ratio) та клас, відповідний до одного з трьох видів: Sensor 1, Sensor 2 або Sensor 3, позначених відповідно 0, 1, 2. Наш алгоритм має приймати чотири властивості одного вимірювального каналу і передбачати, до якого класу (типу датчика) він належить. Наявні в наборі даних мітки можна використовувати для оцінювання якості передбачення.

Для розв'язання завдань кластеризації даних будемо використовувати Python, бібліотеку scikit-learn для завантаження та обробки набору даних і matplotlib для візуалізації. Нижче наведено програмний код для дослідження вихідного набору даних:

```
# Імпортуємо бібліотеки
from sklearn import datasets
import matplotlib.pyplot as plt
# Завантажуємо набір даних
s_df = datasets.load_sensor()
# Методи, доступні для набору даних
print(dir(s_df))
# Признаки
print(s_df.feature_names)
# Мітки
print(s_df.target)
# Імена міток
print(s_df.target_names)
# Розділення набору даних
x_axis = s_df.data[:, 0] # mean
y_axis = s_df.data[:, 1] # variation
# Побудова
plt.xlabel("mean (kPa)")
plt.ylabel("variation (kPa)")
plt.scatter(x_axis, y_axis, c=s_df.target)
plt.show()
```

Результати виконання програми:


```

['DESCR', 'data', 'data_module', 'feature_names', 'filename', 'frame', 'target',
'target_names']
['mean (kPa)', 'variation (kPa)', 'noise level (kPa)', 'signal to noise ratio (dB)']
['Sensor 1' 'Sensor 2' 'Sensor 2']

```

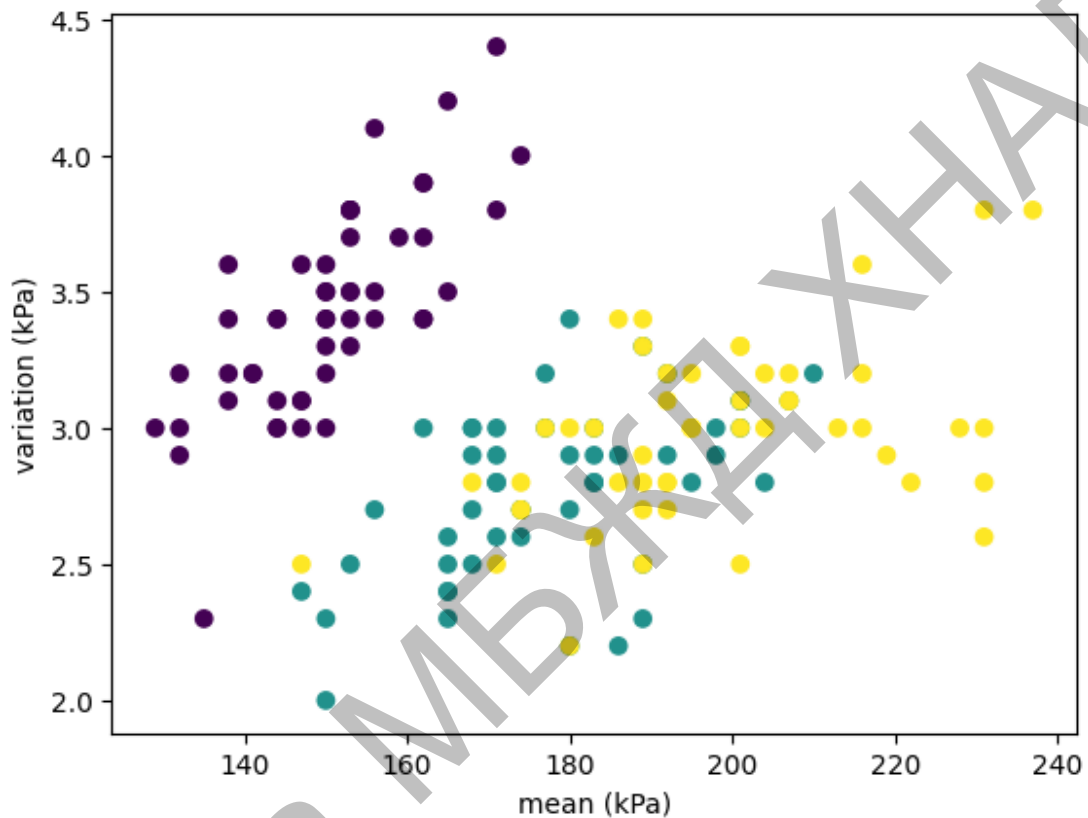


Рисунок 2.2 – Результати кластеризації виміряного тиску датчиків за середнім значенням та дисперсією виміряного тиску

На рис. 2.2 фіолетовим кольором позначено тип датчика тиску Sensor 1, зеленим – Sensor 2 і жовтим - Sensor 2. Під час побудови було взято лише дві ознаки.

Мета кластеризації даних полягає в тому, щоб виділити групи прикладів зі схожими рисами і визначити відповідність прикладів і кластерів. При цьому початково у нас немає прикладів такого розбиття. Це аналогічно тому, як якщо б у наведеному наборі даних у нас не було міток, як на рис. 2.3.

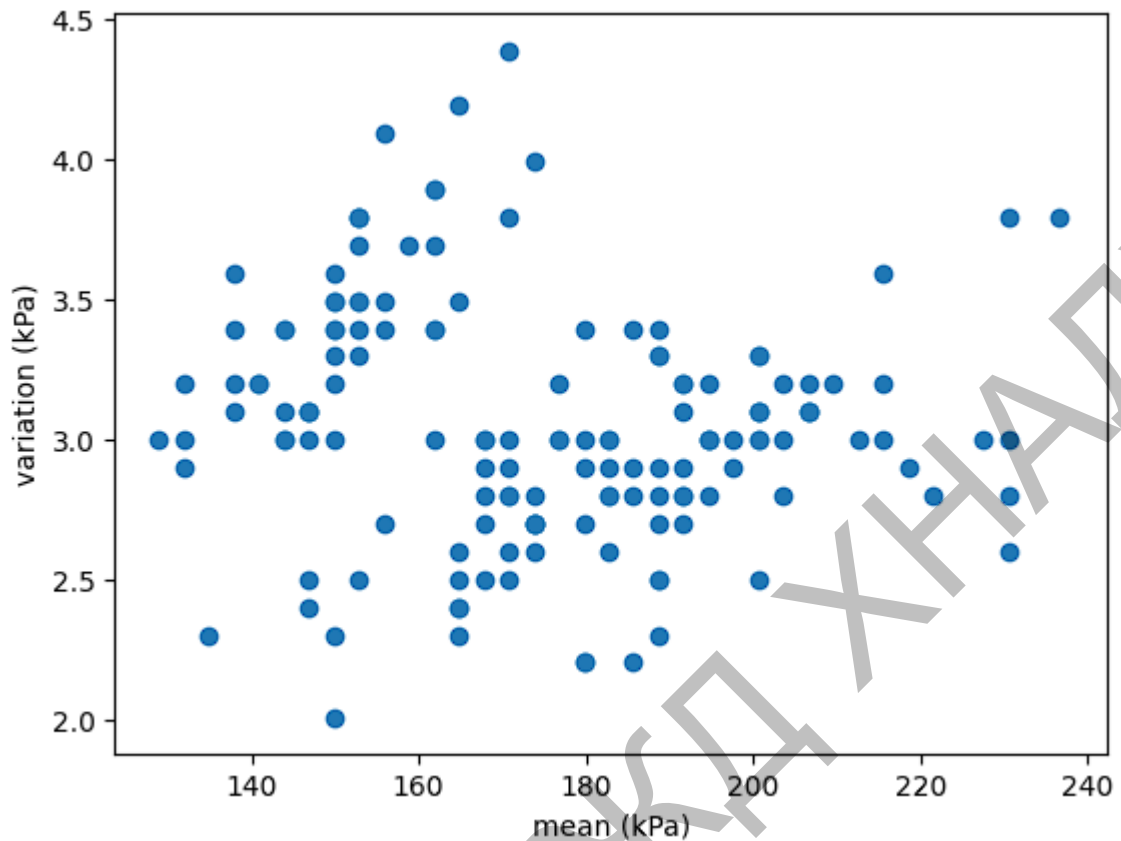


Рисунок 2.3 – Необроблені вхідні дані

Наше завдання – використовуючи всі наявні дані, спрогнозувати відповідність об'єктів вибірки класу та створити кластери.

2.1 Аналіз методу k-середніх

Найпопулярнішим алгоритмом кластеризації даних є метод середніх значень. Це ітераційний алгоритм кластеризації, заснований на мінімізації середньоквадратичного відхилення точок кластерів від центроїдів (середніх координат) цих кластерів.

Спочатку вибирається бажана кількість кластерів. Оскільки нам відомо, що в нашому наборі даних є 3 класи, встановимо параметр моделі `n_clusters` рівний трьом.

Ієрархічна кластеризація. Ієрархічна кластеризація, як випливає з назви, є алгоритмом для створення ієрархії кластерів. Алгоритм починається з зіставлення кожної вибірки даних з власним кластером. Потім два найближчі кластери об'єднуються, і цей процес повторюється, щоб сформувати загальний кластер. Сформовано загальний кластер. Його можна представити у вигляді дендрограми. Ієрархічна кластеризація гірше підходить для кластеризації великих обсягів даних порівняно з методом середніх значень. Це пов'язано з тим, що часова складність алгоритму є лінійною ($O(n)$) для методу k -середніх та квадратичною ($O(n^2)$) для методу ієрархічної кластеризації ($O(n^2)$). У кластеризації за допомогою методу k -середніх алгоритм починає побудову з довільного вибору початкових точок, тому, результати, що генеруються при багаторазовому запуску алгоритму, можуть відрізнятися. Водночас у разі ієрархічної кластеризації результати відтворювані. У той же час, результати відтворюються у випадку ієрархічної кластеризації. Виходячи з форми центру ваги методу k -середніх, метод добре працює, коли кластери мають гіперсферичну форму. Метод k -середніх більш чутливий до зашумлених даних, ніж ієрархічний метод.

2.2 Аналіз методу t-SNE

Метод розподіленого стохастичне розміщення елементів по сусідству (t-SNE) [6, 8,12-16] являє собою один із методів навчання без учителя, використовуваних для візуалізації, наприклад, відображення простору високої розмірності у дво- або тривимірний простір. t-SNE розшифровується як розподілене стохастичне сусіднє вкладення.

Метод моделює кожен об'єкт простору високої розмірності у дво- або трикоординатну точку таким чином, що близькі за характеристиками елементи даних у багатовимірному просторі (наприклад, датасеті з великим числом стовпчиків) проєктують у сусідні точки, а різномірні об'єкти з більшою ймовірністю моделюють точками, що знаходяться далеко одна від одної.

Математичний опис роботи методу можна знайти в [4-7] . Подивимося, як провести моделювання за цим методом за допомогою бібліотеки sklearn. Реалізуємо наступний алгоритм.

```
# Імпорт бібліотек
from sklearn import datasets
from sklearn.manifold import TSNE
import matplotlib.pyplot as plt
# Загрузка датасета
s_df = datasets.load_sensor()
# Визначаємо модель і швидкість навчання
model = TSNE(learning_rate=100)
# Навчаємо модель
transformed = model.fit_transform(s_df.data)
# Представляємо результат у двовимірних координатах
x_axis = transformed[:, 0]*30
y_axis = transformed[:, 1]
plt.xlabel("mean (kPa)")
plt.ylabel("variation (kPa)")
plt.scatter(x_axis, y_axis, c=s_df.target)
plt.show()
```

У цьому разі, як показано на рис. 2.4 кожен екземпляр представлений чотирма координатами – таким чином, при відображенні ознак на площину розмірність простору даних вимірювань тиску знижується з чотирьох до двох.

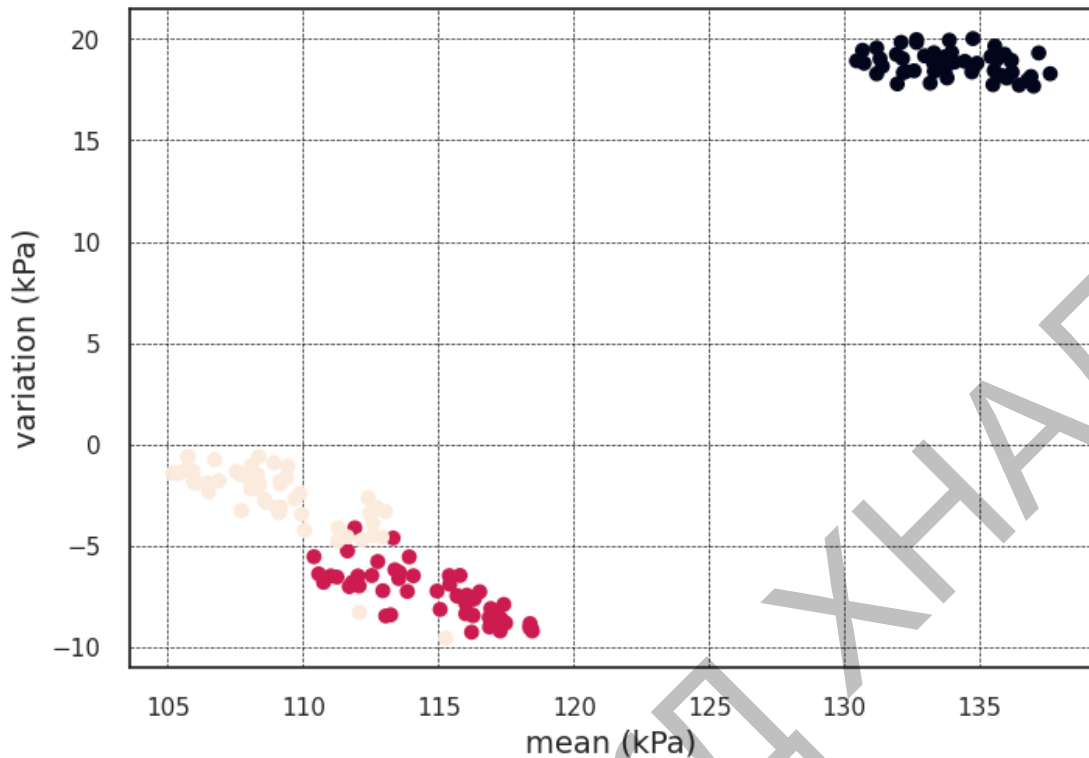


Рисунок 2.4 – Результат кластеризації методом t-SNE

У цьому разі кожен екземпляр представлений чотирма координатами – таким чином, при відображенні ознак на площину розмірність простору знижується з чотирьох до двох.

2.3 Аналіз методу кластеризації на основі щільності ПКЦШ

ПКЦШ – алгоритм кластеризації, який зазвичай використовується в аналізі даних як альтернатива методу k-середніх.

Метод не вимагає попередніх припущень щодо кількості кластерів, але потрібно налаштувати два інших параметри: `eps` і `min_samples`. Ці параметри – це відповідно максимальна відстань між сусідніми точками і мінімальна кількість точок в околиці (кількість сусідів). Про ці екземпляри даних можна сказати, що вони утворюють єдиний кластер. У `scikit-learn` є відповідні значення параметрів за замовчуванням, але, як правило, їх доводиться налаштовувати самостійно.

```

# Імпортуємо бібліотеки
from sklearn.datasets import load_iris
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
# Завантажуємо датасет
iris = load_iris()
# Визначаємо модель
kmeans = KMeans()
# Yfdxf'vj
kmeans.fit(iris.data)
# Зменшуємо розмірність за допомогою методу головних компонент
pca = PCA(n_components=2).fit(iris.data)
pca_2d = pca.transform(iris.data)
# Будуємо відповідно до трьох класів
for i in range(0, pca_2d.shape[0]):
    if kmeans.labels_[i] == 0:
        c1 = plt.scatter(pca_2d[i, 0], pca_2d[i, 1], c='r',
marker='+')
    elif kmeans.labels_[i] == 1:
        c2 = plt.scatter(pca_2d[i, 0], pca_2d[i, 1], c='g',
marker='o')
    elif kmeans.labels_[i] == -1:
        c3 = plt.scatter(pca_2d[i, 0], pca_2d[i, 1], c='b',
marker='*')
plt.legend([c1, c2, c3], ['Кластер 1', 'Кластер 2', 'Шум'])
plt.xlabel("mean (kPa)")
plt.ylabel("variation (kPa)")
plt.show()

```

Результати аналізу кластеризації даних вимірювань тиску в трьох каналній інтелектуальній вимірювальній інформаційній системі методом ПКЩШ [8-12] приведено на рис. 2.5.

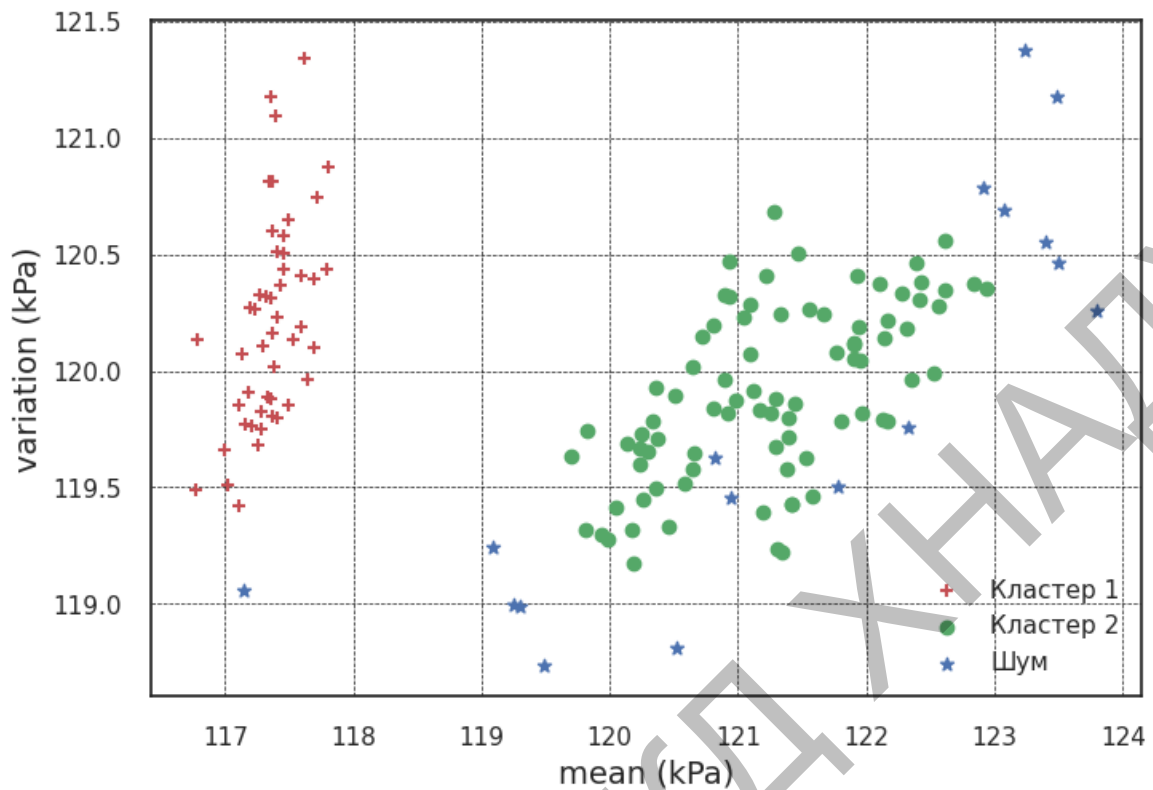


Рисунок 2.5 – Результати роботи алгоритму ПКЦШ

Як показано на рис. 2.5 метод ПКЦШ знайшов два кластери та шум. ПКЦШ відмінно працює на щільних, добре відокремлених один від одного кластерах. При цьому їхня форма абсолютно не важлива. Метод ПКЦШ краще за всіх неспеціалізованих алгоритмів виявляє кластери малої розмірності.

Скупчення різного розміру визначаються непогано (рис.2.6), навіть коли дані сильно зашумлені (рис. 2.7). Тут у ПКЦШ перевага перед k-means, який за просто прийме несиметрично розподілені викиди за кластер (рис. 2.8).

А ось зі скупченнями різної щільності все не так добре. ПКЦШ або знаходить тільки кластер більшої щільності, або все цілком (рис.2.9 – 2.11). Побачити щось підозріле можна хіба що завдяки щасливому випадку, або прогнавши ПКЦШ кілька разів із різними параметрами.

Зі складнішими розподілами також виникають складнощі. Якщо вони добре відокремлені один від одного, ПКЦШ працює. Так на рис. 2.12 (0.4, 5) – чудове розбиття, хоча забагато точок, які алгоритм вважав викидами.

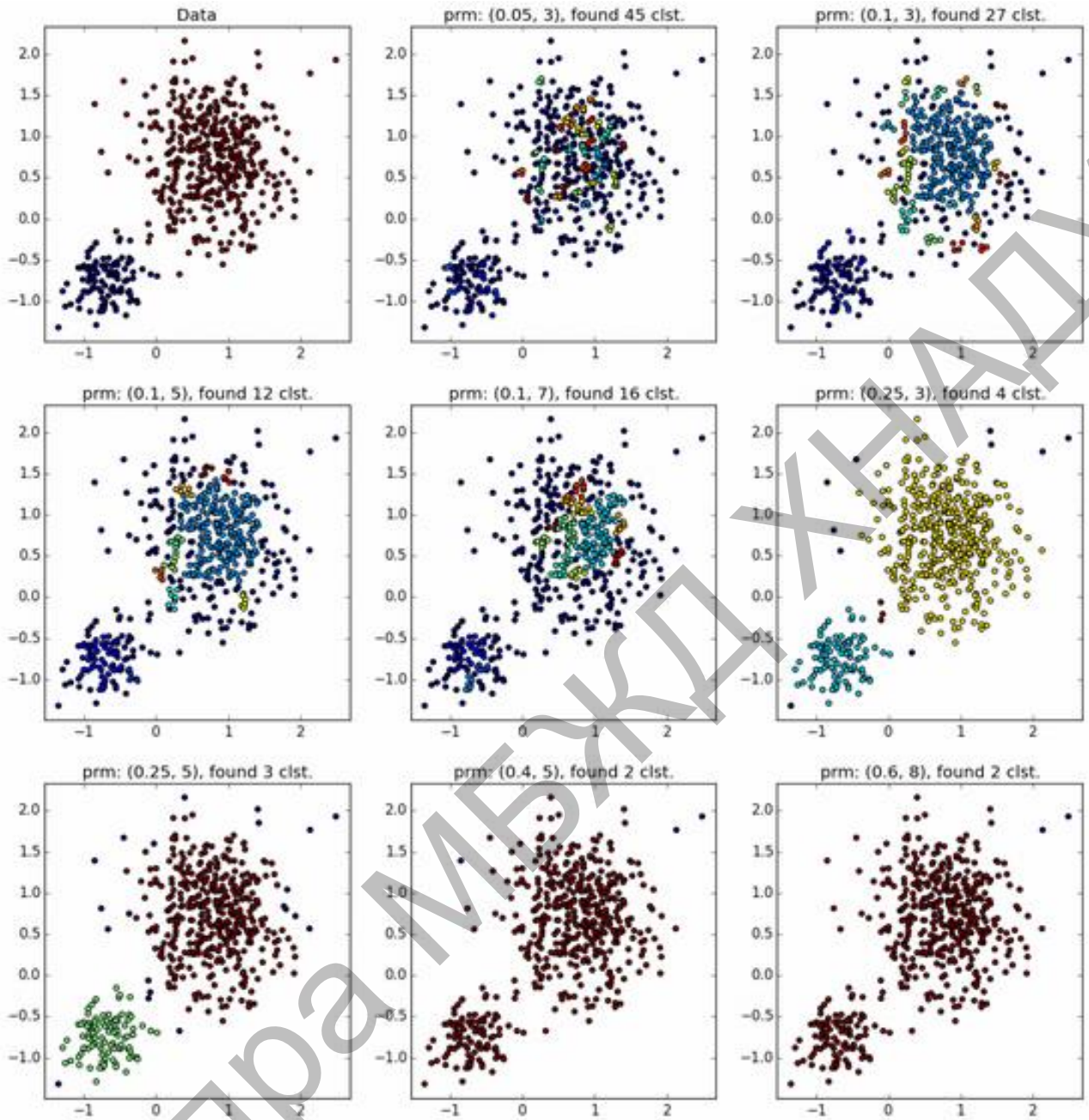


Рисунок 2.6 – Кластеризація скупчення даних вимірювань тиску різного розміру

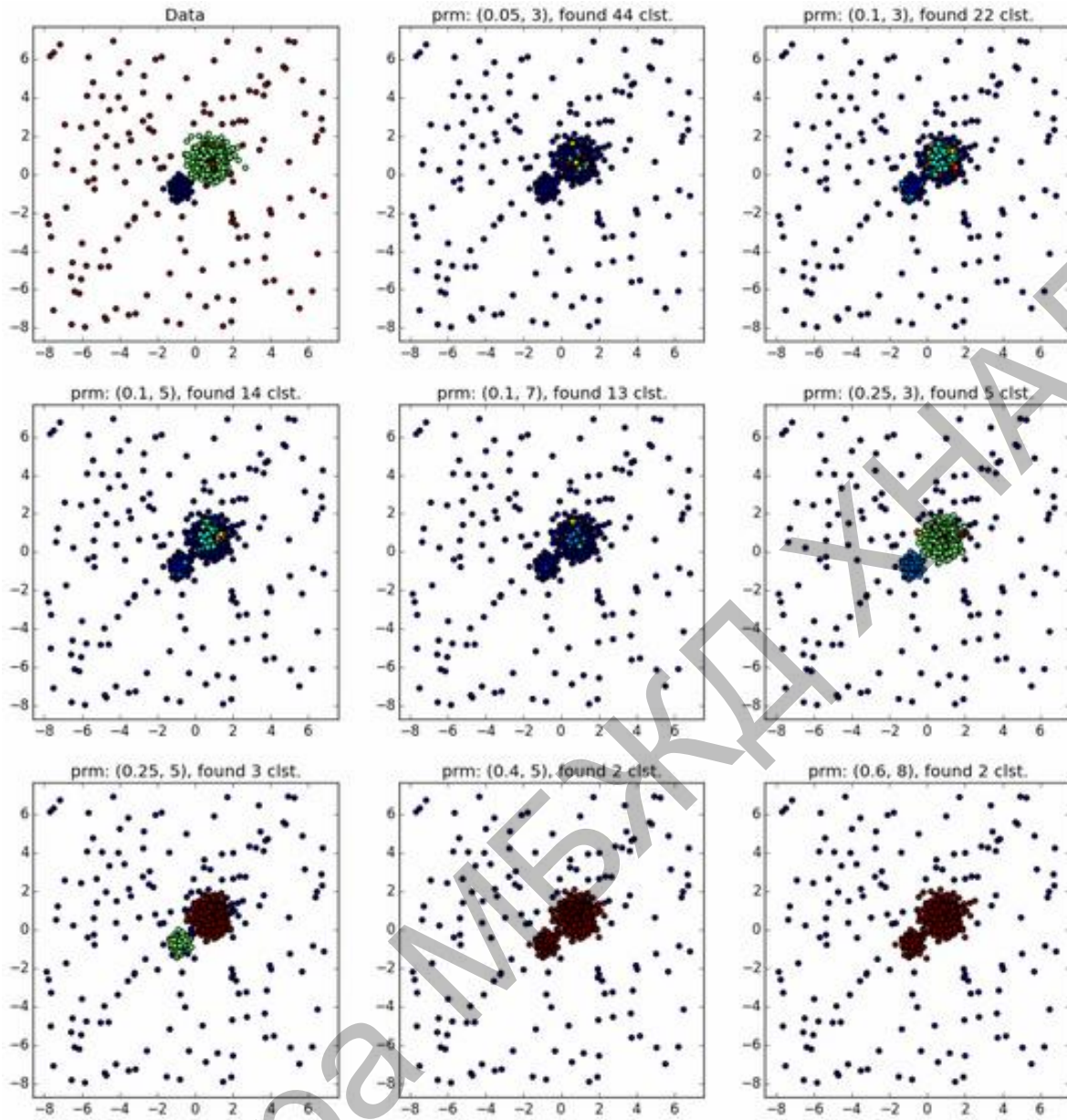


Рисунок 2.7 – Кластеризація сильно зашумлених даних

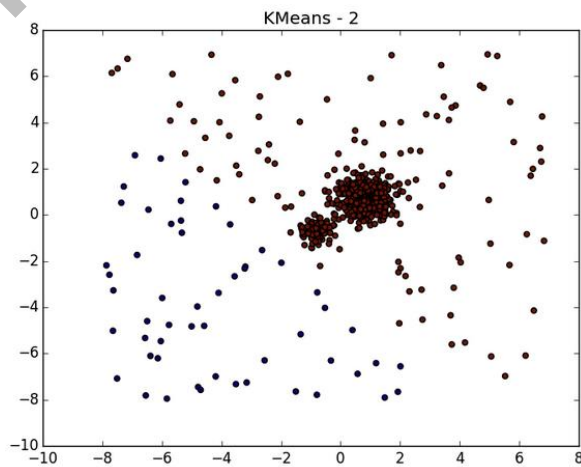


Рисунок 2.8 – Результати кластеризації сильно зашумлених даних методом k-means

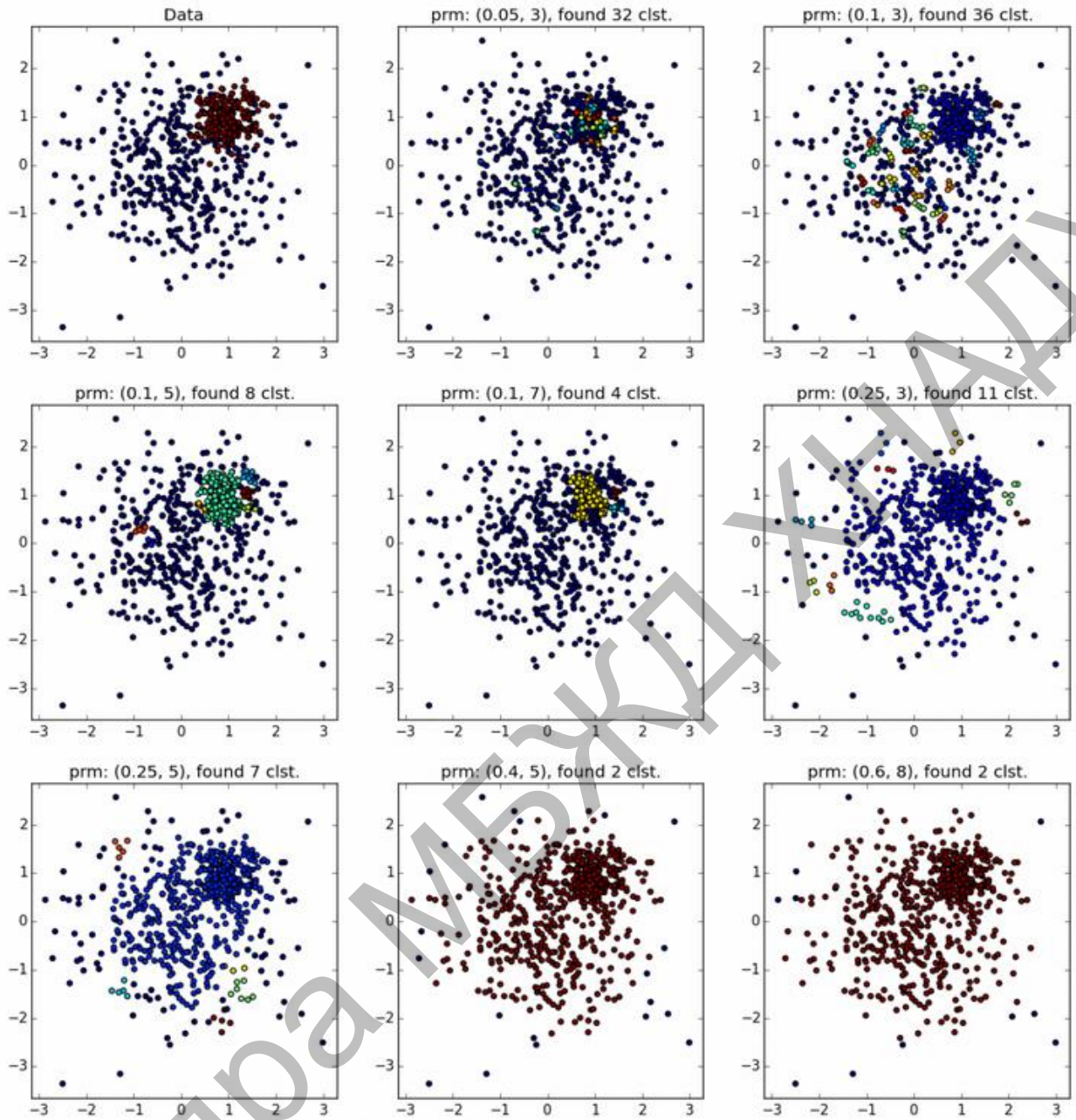


Рисунок 2.9 – Кластеризація скупчення даних вимірювань тиску різної щільності

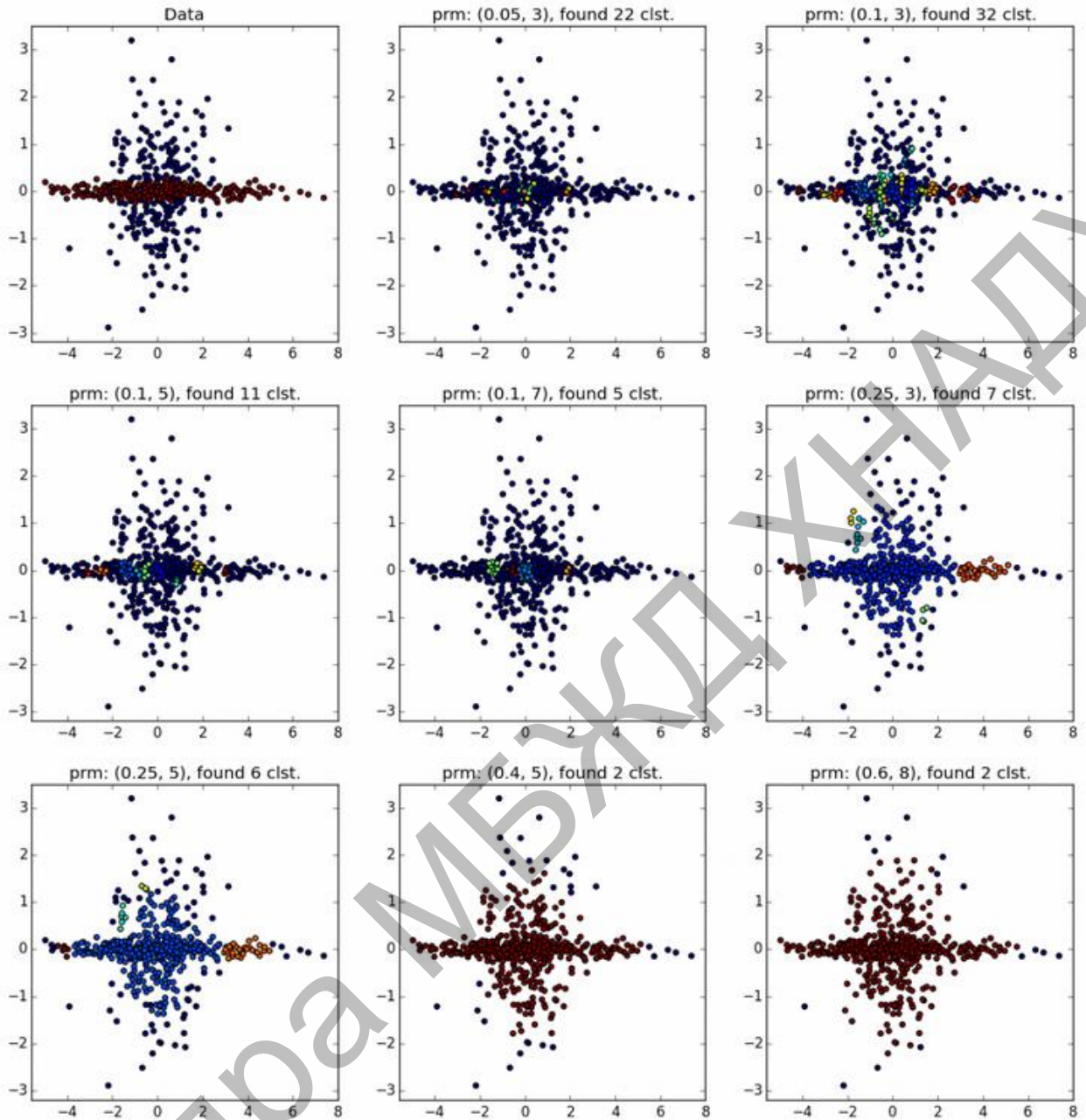


Рисунок 2.10 – Кластеризація скупчення даних вимірювань тиску різної щільності

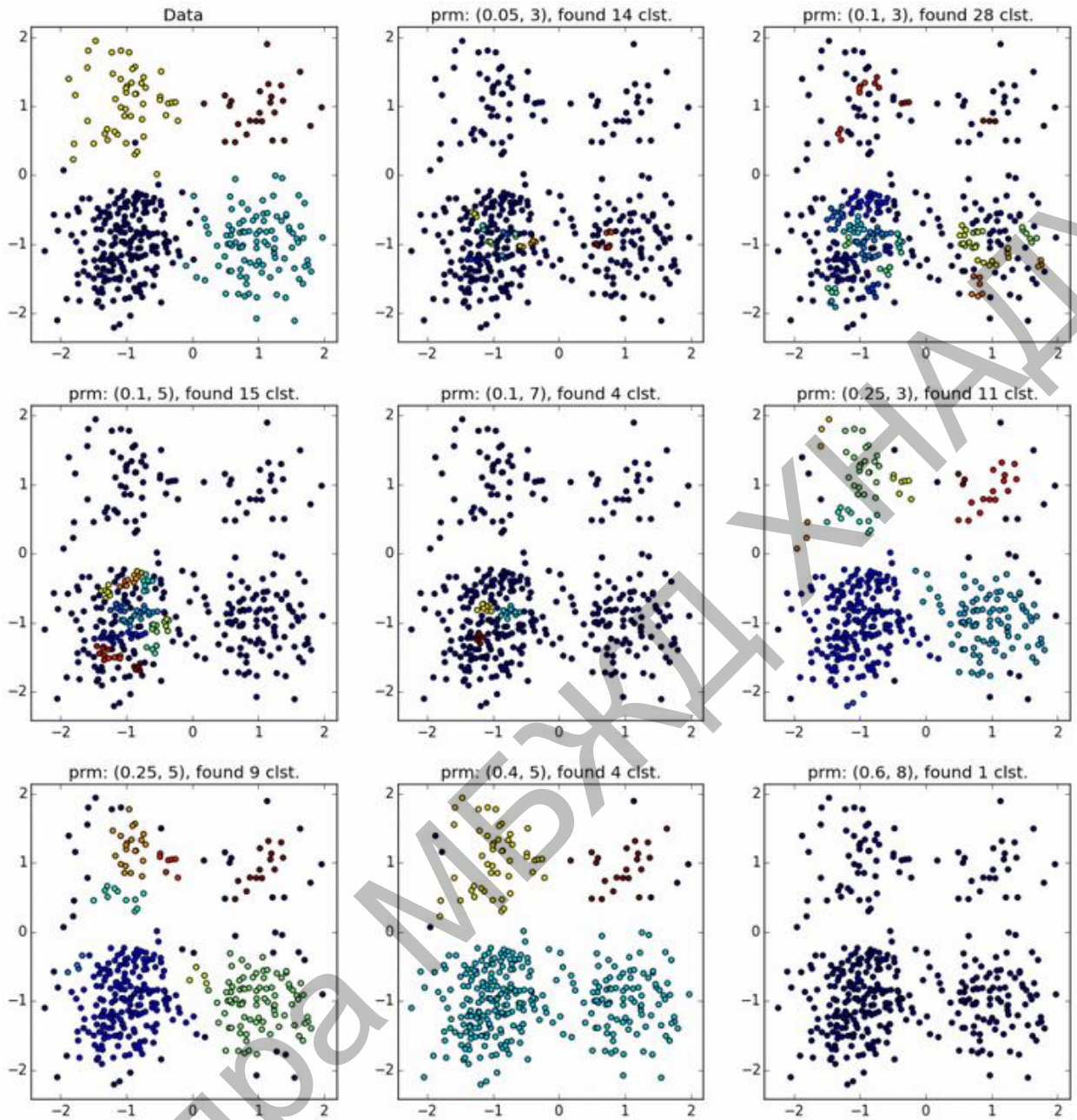


Рисунок 2.11 – Кластеризація скупчення даних вимірювань тиску різної щільності

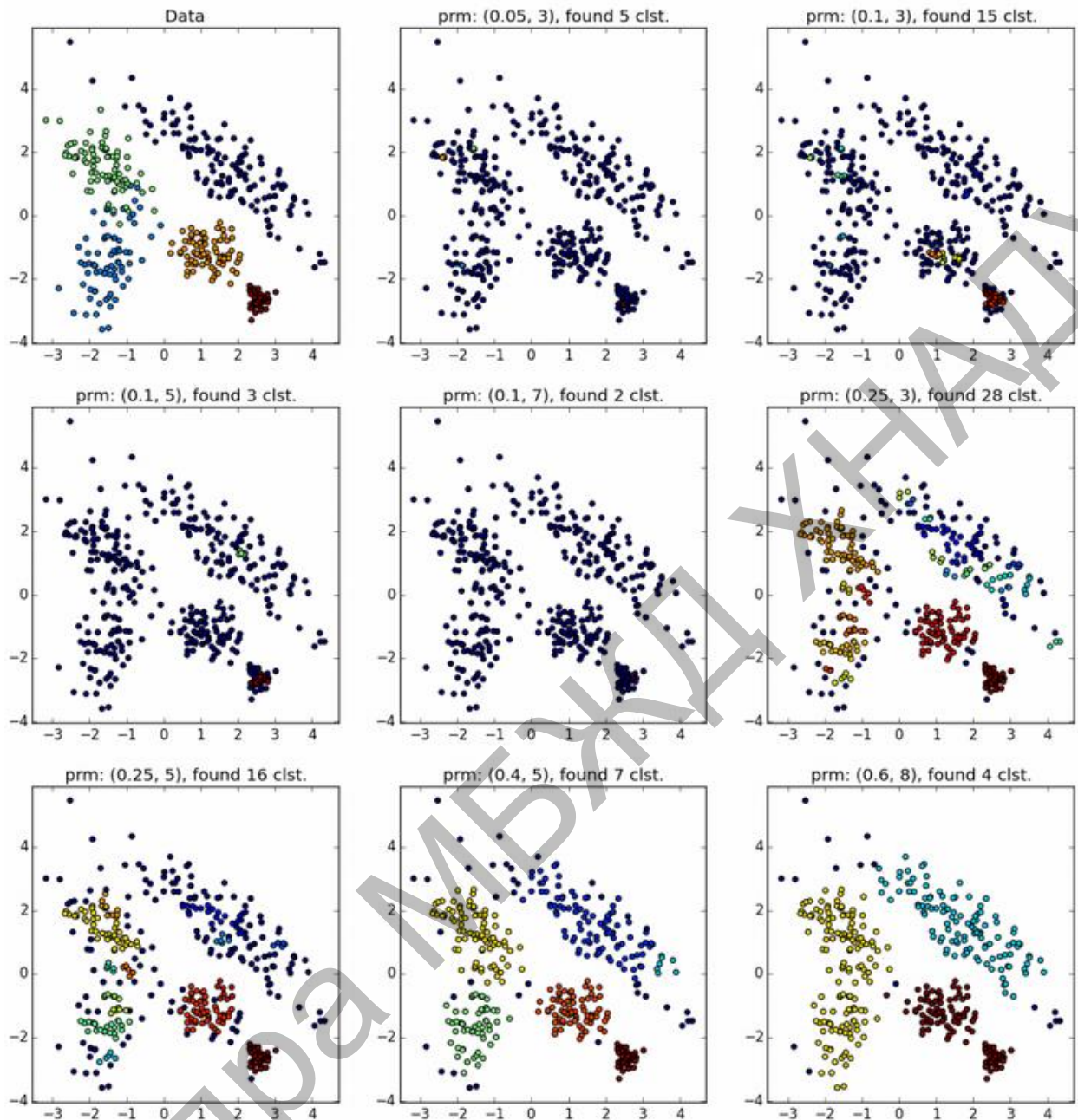


Рисунок 2.12 – Кластеризація масивів даних тиску з різними розподілами

2.4 Висновки до розділу

Похибки кластеризації можуть виникати внаслідок розподілених за різними законами розподілу даних. Деякі з причин цих похибок включають:

1. Неспільність розподілів. Якщо дані в різних кластерах мають різні закони розподілу, алгоритми, які передбачають певні властивості розподілу

(наприклад, гауссівськість), можуть виявитися менш ефективними. Важливо вибирати методи кластеризації, які адаптовані до різних типів розподілу даних.

2. Наявність шуму. Наявність шуму в даних може ускладнювати виявлення чітких кластерів та впливати на роботу алгоритмів кластеризації.

3. Масштабні різниці. Якщо в кожному кластері дані мають значущі різниці в масштабі або дисперсії, це може впливати на ефективність деяких методів, особливо якщо алгоритми використовують відстань між точками.

4. Наявність «важких хвостів». Якщо дані мають тяжкі хвости або важкі хвости (наприклад, експоненціальний або ступеневий розподіл), це може впливати на ефективність алгоритмів, які базуються на припущеннях про нормальний розподіл.

5. Викиди. Наявність викидів у даних може впливати на роботу багатьох алгоритмів кластеризації, особливо тих, які чутливі до шуму та викидів.

Для кращого виявлення кластерів у даних важливо ретельно вивчати їхню природу та враховувати особливості розподілу даних при виборі методів кластеризації. Також може бути корисним використання адаптивних методів або комбінування різних методів для отримання кращих результатів.

3 ДОСЛІДЖЕННЯ ВПЛИВУ ПОХИБОК КЛАСТЕРИЗАЦІЇ ФРЕЙМІВ ДАНИХ ВИМІРЮВАНЬ НА ДОСТОВІРНІСТЬ ПРИЙНЯТТЯ РІШЕНЬ В БАГАТОПАРАМЕТРИЧНИХ ПРОСТОРОВО РОЗПОДІЛЕНИХ ІНТЕЛЕКТУАЛЬНИХ ВИМІРЮВАЛЬНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

Похибки кластеризації в інтелектуальних вимірювальних інформаційних системах можуть значно впливати на достовірність прийняття рішень. Ось кілька способів, якими це може відбуватися:

1. Неправильне групування даних. Якщо алгоритм кластеризації робить помилки і неправильно групує дані, це може призвести до формування неправильних кластерів. В цьому випадку достовірність прийняття рішень залежить від того, наскільки коректно визначені кластери відображають реальну структуру даних вимірювань.

2. Втрата важливої інформації. Похибки кластеризації можуть призвести до втрати важливої інформації про взаємозв'язки між об'єктами або характеристиками. В свою чергу це може вплинути на якість прийняття рішень, оскільки можливість виявлення суттєвих паттернів у даних зменшується.

3. Посилення систематичних помилок. Якщо алгоритм кластеризації систематично допускає певні типи помилок, це може призвести до посилення цих помилок у прийнятті рішень. Систематичні помилки можуть виникнути, наприклад, через неправильний вибір алгоритму, недостатню кількість даних чи некоректно підібрані параметри.

4. Нестійкість до змін в даних. Якщо дані динамічні і змінюються з часом, то неправильності у кластеризації можуть бути особливо проблематичними. Для систем, які базуються на інтелектуальних методах аналізу, важливо, щоб алгоритми були стійкими до змін у вхідних даних.

Всі ці аспекти впливу похибок кластеризації можуть призвести до неправильних аналітичних висновків та, в кінці кінців, сприятимуть прийняттю неточних рішень в інтелектуальних вимірювальних інформаційних системах.

Тому важливо вдосконалювати алгоритми кластеризації та ретельно валідувати їхню ефективність в конкретному контексті.

Похибки кластеризації можуть впливати на різні аспекти інтелектуальної вимірювальної системи [9-11]. Ось деякі з найбільш важливих рішень, які можуть бути вплинуті:

1. Класифікація та розпізнавання образів. Неправильна кластеризація може призвести до неправильної класифікації об'єктів або образів. Це особливо важливо у вимірювальних системах, де важливо точно розпізнавати об'єкти або події.

2. Аналіз аномалій. Якщо система використовує аналіз аномалій для виявлення відхилень або незвичайних подій, невірна кластеризація може призвести до невірної інтерпретації нормальних або аномальних зразків.

3. Прогнозування та вирішення проблем. Якщо система використовує кластеризацію для прогнозування або прийняття рішень, неправильна групування може призвести до неточних прогнозів та невірних рішень.

4. Оптимізація ресурсів. В інтелектуальних вимірювальних системах часто важливо ефективно використовувати ресурси. Неправильна кластеризація може призвести до неефективного використання ресурсів, таких як час, енергія, або обчислювальні потужності.

5. Стабільність та надійність. Кластеризація може впливати на стабільність і надійність системи. Якщо система неправильно розпізнає або групує дані, це може призвести до відмов або погіршення її продуктивності.

6. Взаємодія з користувачем. Якщо інтелектуальна вимірювальна система взаємодіє з користувачем, невірна кластеризація може призвести до неправильних рекомендацій, порад або інформації, що може вплинути на довіру користувача до системи.

Важливо відзначити, що вплив похибок кластеризації залежить від конкретного застосування і вимог системи. Точні наслідки можуть варіюватися в залежності від контексту використання інтелектуальної вимірювальної системи.

Помилки кластеризації в інтелектуальній вимірювальній системі можуть призводити до різних ризиків та негативних наслідків. Ось кілька потенційних ризиків:

1. Неправильні вимірювання та аналіз. Неправильна кластеризація може вплинути на якість вимірювань і аналіз результатів. Якщо об'єкти або події неправильно групуються, може бути складно отримати достовірні та коректні вимірювання.

2. Втрата інформації. При невірній кластеризації інформація може бути втрачена або неправильно інтерпретована. Це може вплинути на можливість виявлення важливих закономірностей або трендів у даних.

3. Низька точність прогнозування. Якщо система використовує кластеризацію для прогнозування, невірні групи можуть призвести до низької точності прогнозів та неправильних висновків.

4. Погіршення якості обслуговування. У просторово розподілених інтелектуальних вимірювальних системах, де якість обслуговування є важливою, неправильна кластеризація може призвести до неправильних відповідей або рекомендацій, що може вплинути на задоволення користувачів.

5. Невірна інтерпретація подій. Кластеризація допомагає групувати події або об'єкти за спільними характеристиками. Невірне групування може призвести до невірної інтерпретації подій та надання неправильних пояснень або причин.

6. Некоректні рішення та дії. Так як інтелектуальна вимірювальна система використовує кластеризацію для прийняття рішень, помилки у кластеризації можуть призвести до неправильних дій або невірних вирішень, що може мати серйозні наслідки.

7. Вплив на безпеку. У деяких випадках невірна кластеризація може призвести до збоїв в безпеці інтелектуальної вимірювальної системи. В ній захист даних вимірювань є дуже значимим фактором.

Для зменшення цих ризиків важливо враховувати особливості даних, вдосконалювати алгоритми кластеризації, проводити аудит та валідацію

результатів, а також враховувати контекст використання інтелектуальної вимірювальної системи.

3.1 Особливості даних вимірюного тиску

Для того щоб правильно вибрати метрику для оцінювання похибок кластеризації масивів вимірюного тиску в багатоканальній просторово розподіленій інтелектуальній вимірювальній системі необхідно вивчити особливості вимірюваного технологічного процесу. В нашому випадку це тиск в системі охолодження атомної електростанції (АЕС) [18]. На рис. 3.1 та рис. 3.2 представлені часові вибірки вимірюного тиску на виході вимірювального каналу тиску в різні часові проміжки доби.

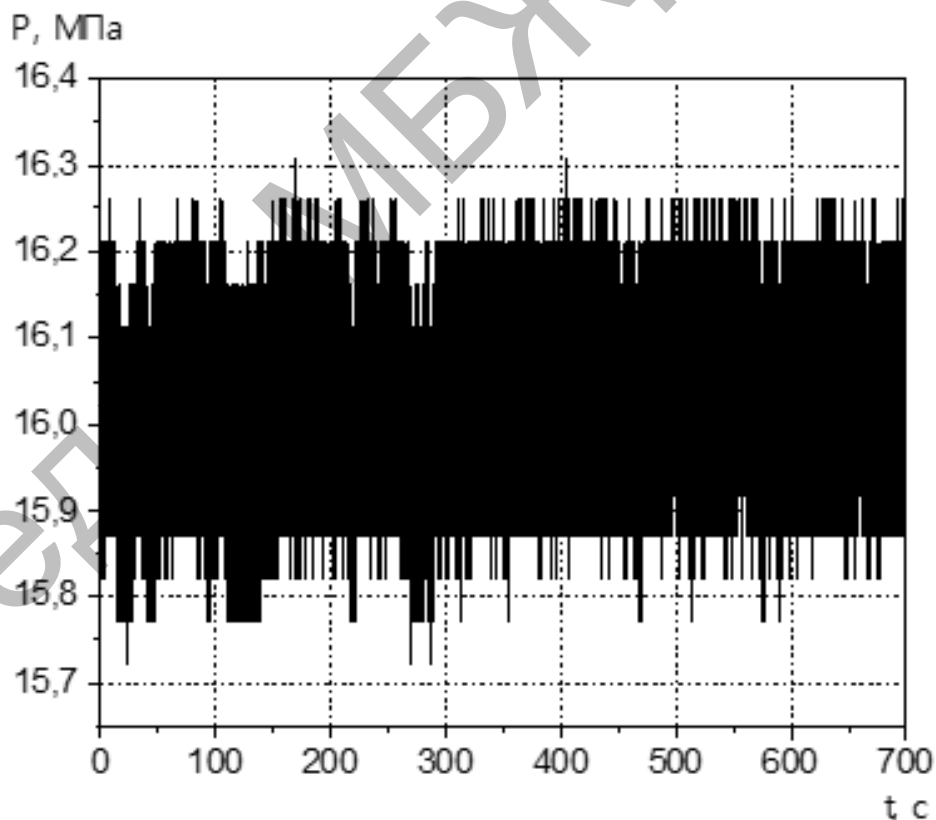


Рисунок 3.1 – Характер вимірюного тиску в нічний час доби

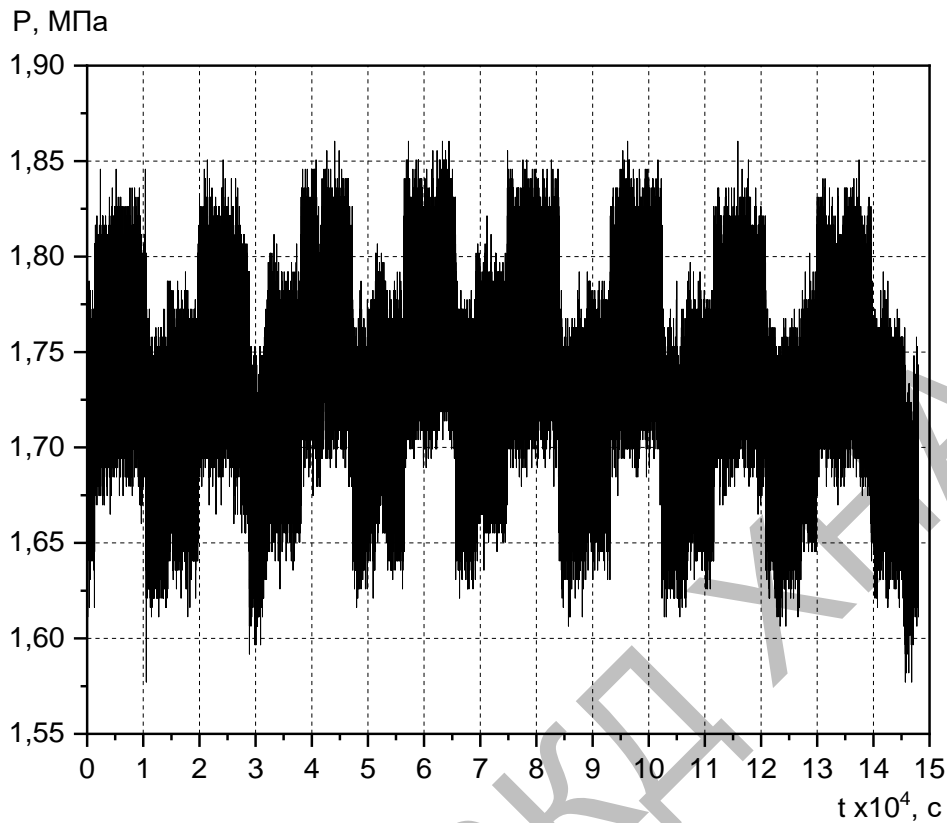


Рисунок 3.2 – Характер виміряного тиску в період з 06.00 до 22.00

Така нестаціонарна поведінка виміряного тиску одним і тим же вимірювальним каналом пояснюється мінливим характером споживаної потужності від АЕС. Ця особливість виміряного тиску – його нестаціонарність в часі вимагає відповідним чином і змінювати метрику, яка використовується для оцінювання похибок кластеризації.

3.1.1 Варіації метрики масивів виміряного тиску

Одиночний, середній і повний зв'язок між виміряними масивами тиску можна використовувати з різними відстанями (або подібностями), зокрема евклідовою відстанню (l_2), манхеттенською відстанню (або *Cityblock*, або l_1), косинусною відстанню або будь-якою попередньо обчисленою матрицею афінності.

Відстань l_1 часто добре підходить для розріджених функцій або розрідженого шуму: тобто багато хто з функцій дорівнює нулю.

Косинусна відстань цікава тим, що вона інваріантна до глобального масштабування сигналу.

Рекомендації щодо вибору метрики – використовувати метрику, яка максимізує відстань між вибірками в різних класах і мінімізує її всередині кожного класу.

Продемонструємо вплив різних метрик на ієрархічну кластеризацію. Для цього будемо використовувати програму розроблену в Python:

```
import matplotlib.pyplot as plt
import numpy as np
from sklearn.cluster import AgglomerativeClustering
from sklearn.metrics import pairwise_distances

np.random.seed(0)
# Формування даних
n_features = 2000
t = np.pi * np.linspace(0, 1, n_features)
def sqr(x):
    return np.sign(np.cos(x))
X = list()
y = list()
for i, (phi, a) in enumerate([(0.5, .15), (.5, .6), (.3, .2)]):
    for _ in range(30):
        phase_noise = .01 * np.random.normal()
        amplitude_noise = .04 * np.random.normal()
        additional_noise = 1 - 2 * np.random.rand(n_features)
        # розріджений шум
        additional_noise[np.abs(additional_noise) < .997] = 0
        X.append(12 * ((a + amplitude_noise)
                       * (sqr(6 * (t + phi + phase_noise)))
                       + additional_noise))
        y.append(i)
X = np.array(X)
y = np.array(y)
n_clusters = 3
labels = ('Sensor1', 'Sensor2', 'Sensor3')
# схема маркування правдивої інформації
```

```

plt.figure()
plt.axes([0, 0, 1, 1])
for l, c, n in zip(range(n_clusters), 'rgb', labels):
    lines = plt.plot(X[y == l].T, c=c, alpha=.5)
    lines[0].set_label(n)
plt.legend(loc='best')
plt.axis('tight')
plt.axis('off')
plt.suptitle("Проста істина", size=14)
# графік відстаней
for index, metric in enumerate(["cosine", "euclidean", "cityblock"]):
    avg_dist = np.zeros((n_clusters, n_clusters))
    plt.figure(figsize=(5, 4.5))
    for i in range(n_clusters):
        for j in range(n_clusters):
            avg_dist[i, j] = pairwise_distances(X[y == i], X[y == j],
                                                metric=metric).mean()
    avg_dist /= avg_dist.max()
    for i in range(n_clusters):
        for j in range(n_clusters):
            plt.text(i, j, '%5.3f' % avg_dist[i, j],
                    verticalalignment='center',
                    horizontalalignment='center')
    plt.imshow(avg_dist, interpolation='nearest', cmap=plt.cm.gnuplot2,
               vmin=0)
    plt.xticks(range(n_clusters), labels, rotation=45)
    plt.yticks(range(n_clusters), labels)
    plt.colorbar()
    plt.suptitle("Міжкласні %s відстані" % metric, size=14)
    plt.tight_layout()
# Результати кластеризації масивів
for index, metric in enumerate(["cosine", "euclidean", "cityblock"]):
    model = AgglomerativeClustering(n_clusters=n_clusters,
                                    linkage="average", affinity=metric)
    model.fit(X)
    plt.figure()
    plt.axes([0, 0, 1, 1])
    for l, c in zip(np.arange(model.n_clusters), 'rgbk'):
        plt.plot(X[model.labels_ == l].T, c=c, alpha=.5)
    plt.axis('tight')
    plt.axis('off')
    plt.suptitle("Агломеративна кластеризація (спорідненість=%s)"
                % metric, size=14)
plt.show()

```

Програма розроблена для того, щоб показати ефект від вибору різних метрик. Його застосовано до масивів даних вимірювань, які можна розглядати як вектор високої розмірності. Дійсно, різниця між метриками зазвичай більш виражена у високій розмірності (зокрема, для евклідової та cityblock).

Ми генеруємо дані з трьох масивів вимірювань даних. Дві з них (масив 1 і масив 2) пропорційні одна до одної. Косинусна відстань є інваріантною до масштабування даних, тому вона не може розрізнити ці два масиви. Таким чином, навіть за відсутності шуму, кластеризація з використанням цієї відстані не розділить масиви даних вимірювань тиску 1 і 2.

Ми додаємо до цих осцилограм шум спостереження. Ми генеруємо дуже розріджений шум: лише 6% часових точок містять шум. В результаті, норма $l1$ цього шуму (тобто cityblock відстань) набагато менша, ніж норма $l2$ ("евклідова" відстань). Це можна побачити на матрицях міжкласових відстаней (рис. 3.3, 3.4, 3.5): значення по діагоналі, які характеризують поширення класу, набагато більші для евклідової відстані, ніж для відстані cityblock.

Коли ми застосовуємо кластеризацію до даних, ми бачимо, що кластеризація відображає те, що було в матрицях відстаней. Дійсно, для евклідової відстані (рис. 3.6) класи погано розділені через шум, і тому кластеризація не розділяє масиви вимірювань. Для відстані cityblock (рис. 3.7) поділ є добрим, і класи масивів вимірювань відновлюються. Нарешті, косинусоїдальна відстань (рис. 3.8) взагалі не розділяє масиви даних вимірювань 1 і 2, тому кластеризація відносить їх до одного кластера.

Міжкласні cosine відстані

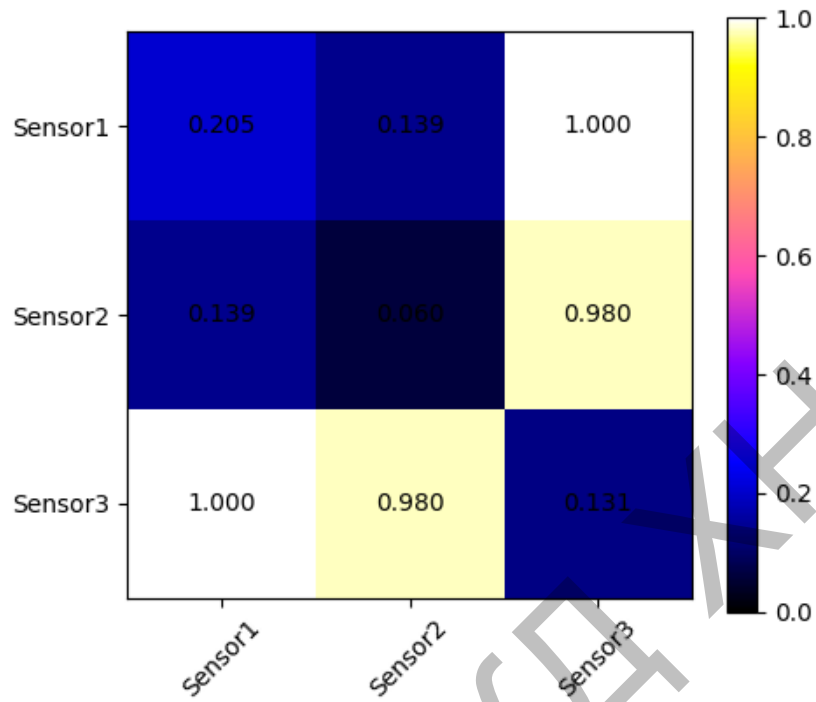


Рисунок 3.3 – Міжкласні косинусні відстані

Міжкласні euclidean відстані

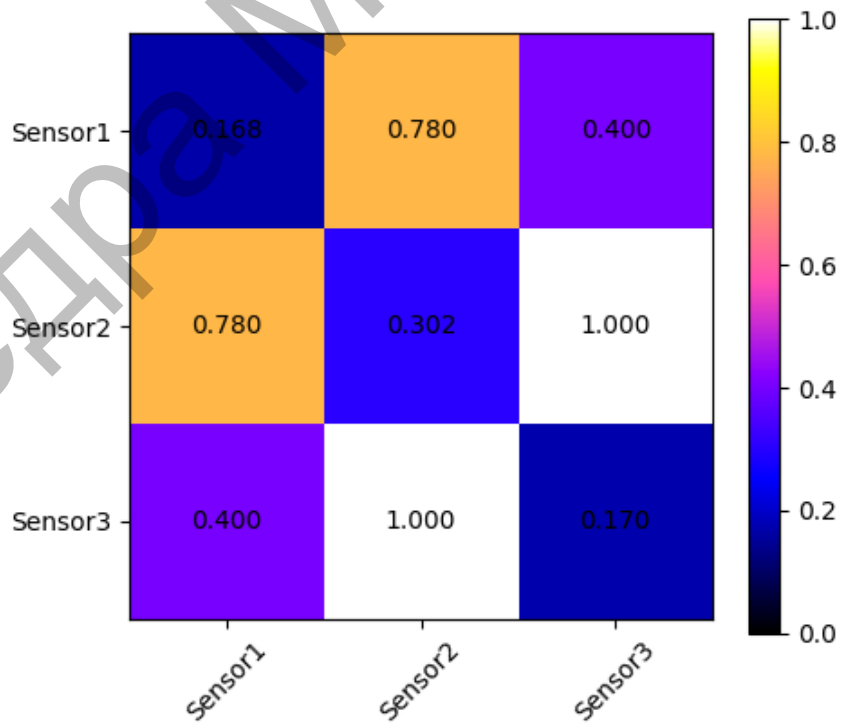


Рисунок 3.4 – Міжкласні евклідові відстані

Міжкласні cityblock відстані

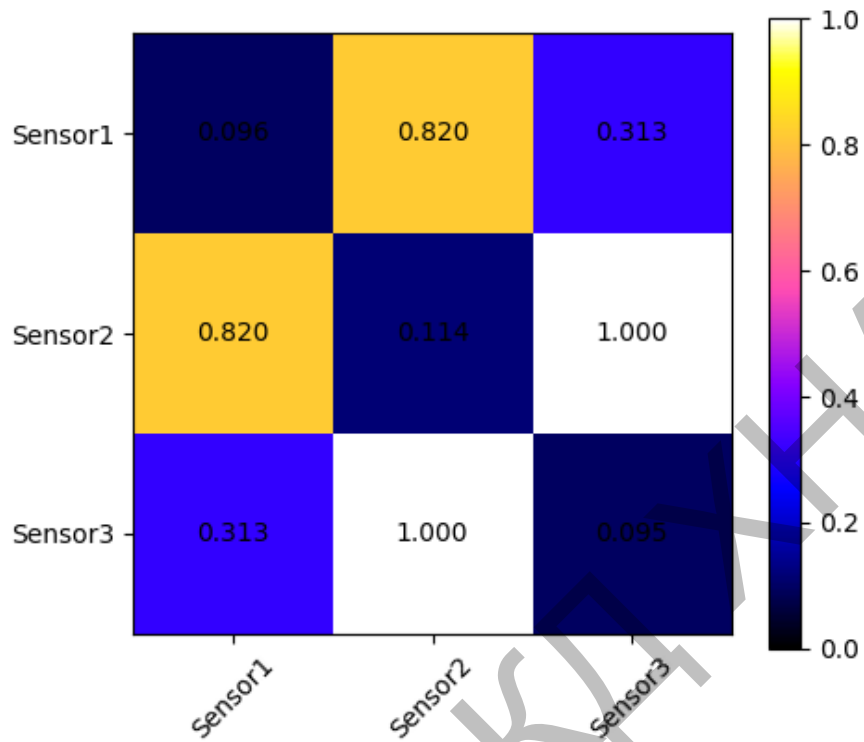


Рисунок 3.5 – Міжкласні cityblock відстані

Агломеративна кластеризація (спорідненість=cosine)

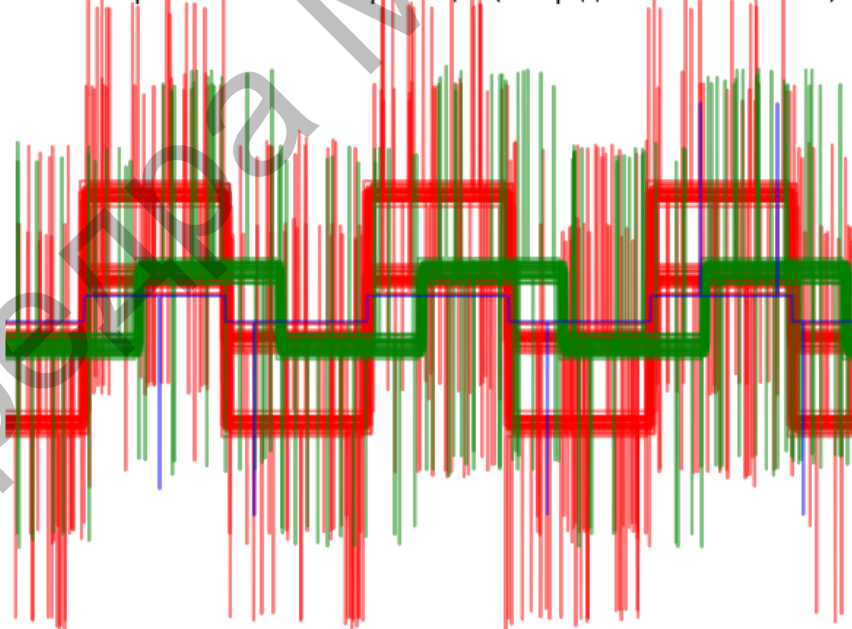


Рисунок 3.6 – Англомеративна кластеризація (cosine спорідненість)

Агломеративна кластеризація (спорідненість=euclidean)

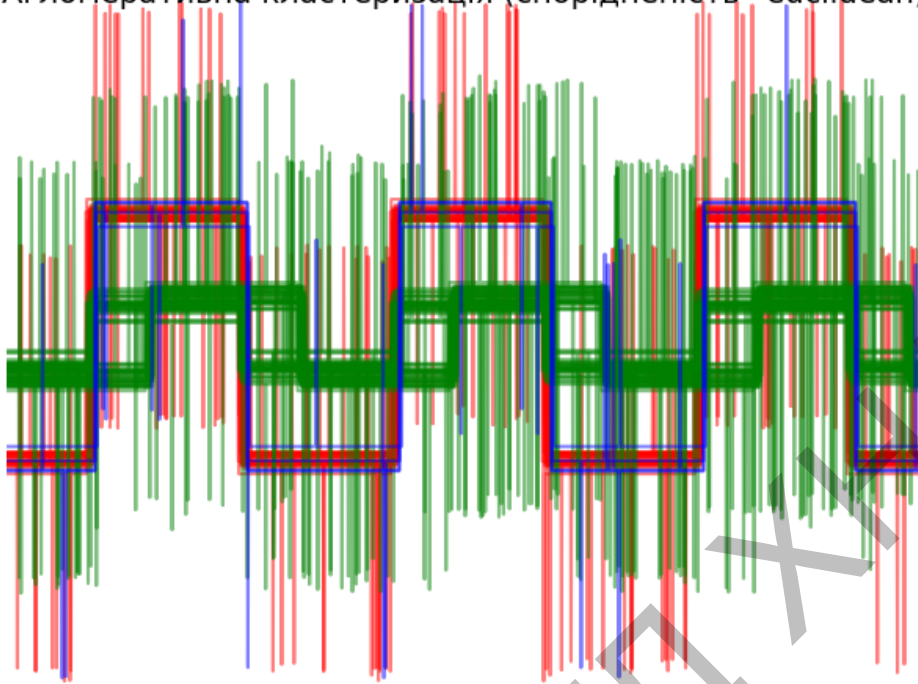


Рисунок 3.7 – Англомеративна кластеризація (euclidean спорідненість)

Агломеративна кластеризація (спорідненість=cityblock)

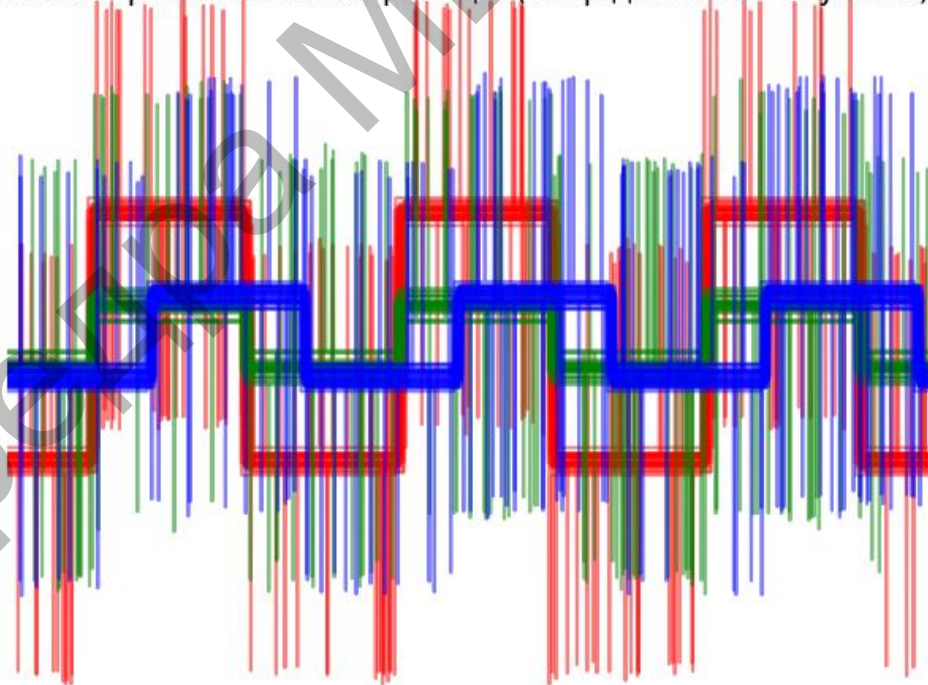


Рисунок 3.8 – Англомеративна кластеризація (cityblock спорідненість)

3.2 Метрики оцінювання похибок кластеризації

Оцінка якості результатів кластеризації може використовувати різні метрики в залежності від конкретних завдань і характеристик даних. Деякі з загальноприйнятих метрик включають:

1. Adjusted Rand Index (ARI): Ця метрика вимірює подібність між фактичною і передбаченою кластеризацією, кориговану на випадковий вибір. ARI повертає значення від -1 до 1, де 1 означає ідеальну узгодженість.

2. Silhouette Score: Цей показник оцінює, наскільки об'єкти в одному кластері схожі між собою, порівняно з об'єктами в інших кластерах. Значення знаходиться в діапазоні від -1 до 1, де високий показник вказує на добру якість кластеризації.

3. Homogeneity, Completeness, V-measure: Ці метрики використовуються для вимірювання рівня однорідності (Homogeneity), повноти (Completeness) і їх зваженої середньої (V-measure) кластеризації.

4. Davies-Bouldin Index: Цей індекс використовує внутрішні та зовнішні відстані між кластерами для оцінки якості кластеризації. Менше значення вказує на кращу якість.

5. Calinski-Harabasz Index: Цей індекс вимірює відношення між внутрішньою та зовнішньою дисперсією для оцінки якості кластеризації. Велике значення вказує на кращу якість.

6. Fowlkes-Mallows Index: Ця метрика використовується для вимірювання схожості між фактичною і передбаченою кластеризацією. Якщо ви розглядаєте задачу бінарної класифікації, цей індекс може бути корисним.

Важливо враховувати, що немає універсальної метрики, яка підходить для всіх типів даних та завдань кластеризації. Вибір метрики повинен бути здійснений з урахуванням конкретного контексту дослідження.

Для оцінки похибок кластеризації можна використовувати метрики, такі як Adjusted Rand Index (ARI) або Silhouette Score. Вам знадобиться бібліотека scikit-

learn для використання цих метрик. Якщо вона не встановлена, ви можете її встановити за допомогою команди `pip install scikit-learn`.

Для оцінки похибок кластеризації була використана програма на Python, яка використовує KMeans для кластеризації і виводить значення ARI:

```
from sklearn.cluster import KMeans
from sklearn.metrics import adjusted_rand_score,
silhouette_score
from sklearn.datasets import make_blobs
# Генерація випадкових даних для тестування
X, y = make_blobs(n_samples=300, centers=3, random_state=42)
# Визначення кількості кластерів (у реальному випадку це може
бути невідомо)
n_clusters = 3
# Модель KMeans
kmeans = KMeans(n_clusters=n_clusters, random_state=42)
predicted_labels = kmeans.fit_predict(X)
# Оцінка похибок кластеризації
ari = adjusted_rand_score(y, predicted_labels)
silhouette = silhouette_score(X, predicted_labels)
# Виведення результатів
print(f"Adjusted Rand Index (ARI): {ari}")
print(f"Silhouette Score: {silhouette}")
```

Ця програма використовує `make_blobs` з `scikit-learn` для генерації випадкових даних з трьома кластерами. В реальному випадку вам слід замінити `X` та `y` на ваші власні дані. Потім вона використовує KMeans для кластеризації та обчислює ARI та Silhouette Score для оцінки якості кластеризації.

Проведені дослідження показали, що ці метрики можуть бути чутливими до форми та розміру кластерів.

3.3 Однорідність, повнота і V-міра

Знаючи про призначення основних класів істинності вибіркам, можна визначити деяку інтуїтивно зрозумілу метрику, використовуючи умовний ентропійний аналіз. Визначимо такі дві цілі для нашої задачі кластерного аналізу:

- однорідність, кожен кластер містить тільки членів одного класу;
- повнота, усі члени даного класу належать до одного кластера.

Ми можемо перетворити ці концепції на партитури `homogeneity_score` і `completeness_score`. Обидва обмежені знизу 0,0 і вище 1,0 (що вище, то краще):

```
from sklearn import metrics
labels_true = [0, 0, 0, 1, 1, 1]
labels_pred = [0, 0, 1, 1, 2, 2]
metrics.homogeneity_score(labels_true, labels_pred)
0.6666666666666669
metrics.completeness_score(labels_true, labels_pred)
0.420619835714305
```

Їхнє гармонійне середнє значення, зване V-мірою, обчислюється за функцією `v_measure_score`:

```
metrics.v_measure_score(labels_true, labels_pred)
0.5158037429793889
```

Ця функція визначається згідно з наступним математичним виразом [9]:

$$v = \frac{(1 + \beta) \times \text{homogeneity} \times \text{completeness}}{(\beta \times \text{homogeneity} + \text{completeness})}. \quad (3.1)$$

За замовчуванням $\beta = 1$, але для використання значення менше ніж 1 використання функції дещо зміниться:

```
metrics.v_measure_score(labels_true, labels_pred, beta=0.6)
0.5467344787062375
```

Більшу вагу буде приписано однорідності, а використання значення більше

1:

```
metrics.v_measure_score(labels_true, labels_pred, beta=1.8)
0.48447946234141737
```

більшу вагу буде приписано повноті.

V-міра фактично еквівалентна взаємній інформації (NMI), при цьому функція агрегування є середнім арифметичним [B2011].

Однорідність, повноту і V-міру можна обчислити одразу, використовуючи `homogeneity_completeness_v_measure` наступним фрагментом коду Python:

```
metrics.homogeneity_completeness_v_measure(labels_true,
labels_pred)
(0.6666666666666669, 0.420619835714305, 0.5158037429793889)
```

Переваги використання V-міри:

— Обмежена кількість балів: 0.0 - найгірший результат, 1.0 - відмінний результат.

— Інтуїтивна інтерпретація: кластеризація з низькою V-мірою. Можна якісно проаналізувати з погляду однорідності та повноти, щоб краще зрозуміти, яких "помилко" припускаються під час завдання.

— Оскільки не робиться жодних припущень щодо структури кластерів, алгоритми кластеризації, такі як k-середні, які припускають ізотропну форму крапель, можна використовувати для порівняння результатів з алгоритмами спектральної кластеризації, які можуть знаходити кластери зі "складеною" формою.

Наступні графіки, які приведені на рис. 3.9 та на рис. 3.10 демонструють вплив кількості кластерів та кількості вибірок на різні метрики оцінки ефективності кластеризації.

Не скориговані міри, такі як V-міра, показують залежність між кількістю кластерів і кількістю зразків: середнє значення V-міри випадкового маркування значно зростає, коли кількість кластерів наближається до загальної кількості зразків, використаних для обчислення міри.

Скоригована на випадковість міра, така як ARI, показує деякі випадкові варіації, зосереджені навколо середньої оцінки 0,0 для будь-якої кількості зразків і кластерів.

Отже, тільки скориговані міри можна безпечно використовувати як індекс консенсусу для оцінки середньої стабільності алгоритмів кластеризації для заданого значення k на різних підвбірках набору даних, що перекриваються.

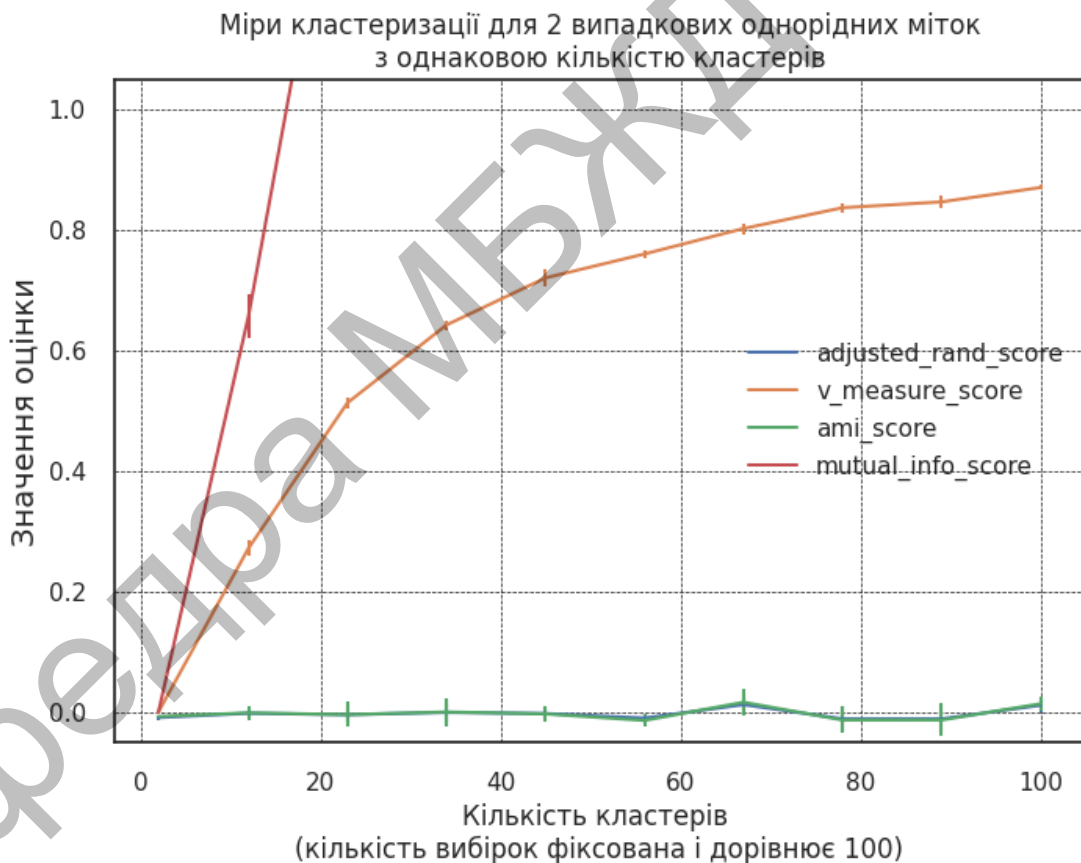


Рисунок 3.9 – Міри кластеризації для двох однорідних масивів виміряного тиску з однаковою кількістю кластерів

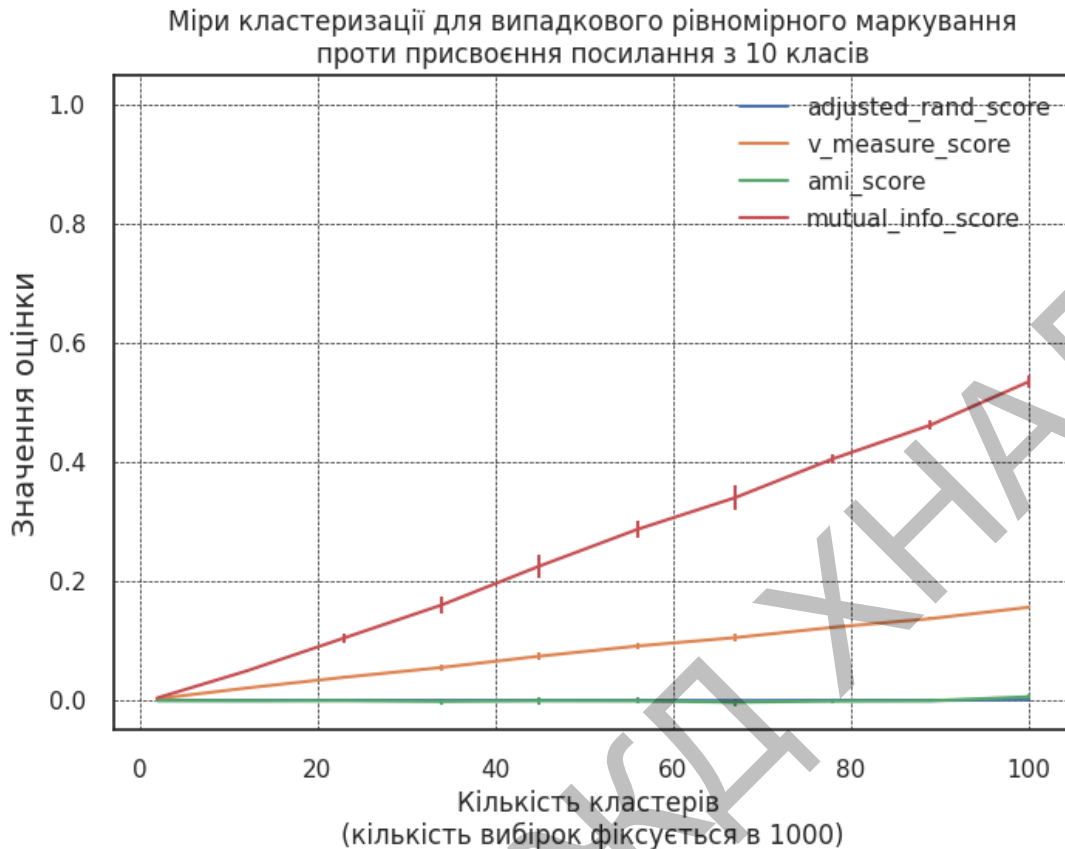


Рисунок 3.10 – Міри кластеризації для двох випадкових масивів вимірюного тиску з однаковою кількістю кластерів

3.4 Оцінювання впливу похибок кластеризації на достовірність прийняття рішення

Достовірність прийняття рішень в інтелектуальній вимірювальній інформаційній системі (ІВІС) зазвичай визначається рядом факторів. Ось деякі з них:

1. Якість вхідних даних. Достовірність рішень залежить від якості та достовірності вхідної інформації. Якщо вхідні дані неточні або застарілі, це може призвести до невірних висновків.

2. Алгоритми та моделі. Використання правильних алгоритмів та моделей грає важливу роль в достовірності рішень. Якщо алгоритми погано підібрані або моделі неправильно навчені, це може вплинути на точність прогнозів та рішень.

3. Контроль якості даних. Впровадження систем контролю якості даних допомагає виявляти та виправляти помилки в інформації, що використовується для прийняття рішень.

4. Моніторинг та звітність. Забезпечення системи моніторингу та звітності дозволяє вчасно виявляти проблеми та втручатися для їх вирішення.

5. Експертна оцінка. Залучення експертів для оцінки та перевірки рішень може допомогти підтвердити їх достовірність.

6. Прозорість системи. Якщо система прозора і розуміється користувачем, вона надає можливість більш ефективного контролю та розуміння того, як саме приймаються рішення.

7. Врахування не визначеності. Інтелектуальні системи повинні бути здатні враховувати та керувати не визначеністю вхідних даних та моделей.

Ці аспекти спільно впливають на достовірність рішень в інтелектуальних вимірювальних інформаційних системах, і їх ефективна інтеграція може підвищити точність та надійність прийнятих рішень.

Для оцінювання впливу похибок кластеризації на достовірність прийняття рішення можна використати метрики якості кластеризації та подальші аналізи. Однією з таких метрик є Adjusted Rand Index (ARI). Нижче наведено програму на Python, яка використовує бібліотеку scikit-learn для виконання кластеризації та розрахунку ARI:

```
from sklearn.cluster import KMeans
from sklearn import datasets
from sklearn.metrics import adjusted_rand_score
import numpy as np
# Згенеруємо деякі зразки для прикладу (можна замінити на свої дані)
data, true_labels = datasets.make_blobs(n_samples=300, centers=3,
random_state=42)
# Введемо похибку в кластеризацію (моделюємо ситуацію помилкового
кластерування)
random_indices = np.random.choice(len(true_labels), size=30,
replace=False)
```

```

data[random_indices] = np.random.uniform(low=-10, high=10,
size=(30, 2))
# Виконаємо кластеризацію (замість KMeans можна використовувати
інші алгоритми)
kmeans = KMeans(n_clusters=3, random_state=42)
predicted_labels = kmeans.fit_predict(data)
# Розрахунок Adjusted Rand Index (ARI)
ari = adjusted_rand_score(true_labels, predicted_labels)
print(f"Adjusted Rand Index: {ari}")
# Аналіз впливу похибок кластеризації на достовірність прийняття
рішення
if ari > 0.8:
    print("Дуже добре: Кластеризація добре відповідає справжнім
кластерам.")
elif ari > 0.5:
    print("Добре: Кластеризація має середню схожість із справжніми
кластерами.")
else:
    print("Потрібно перевірити: Кластеризація може бути
недостовірною.")
Adjusted Rand Index: 0.8366144125970839
Дуже добре: Кластеризація добре відповідає справжнім кластерам.

```

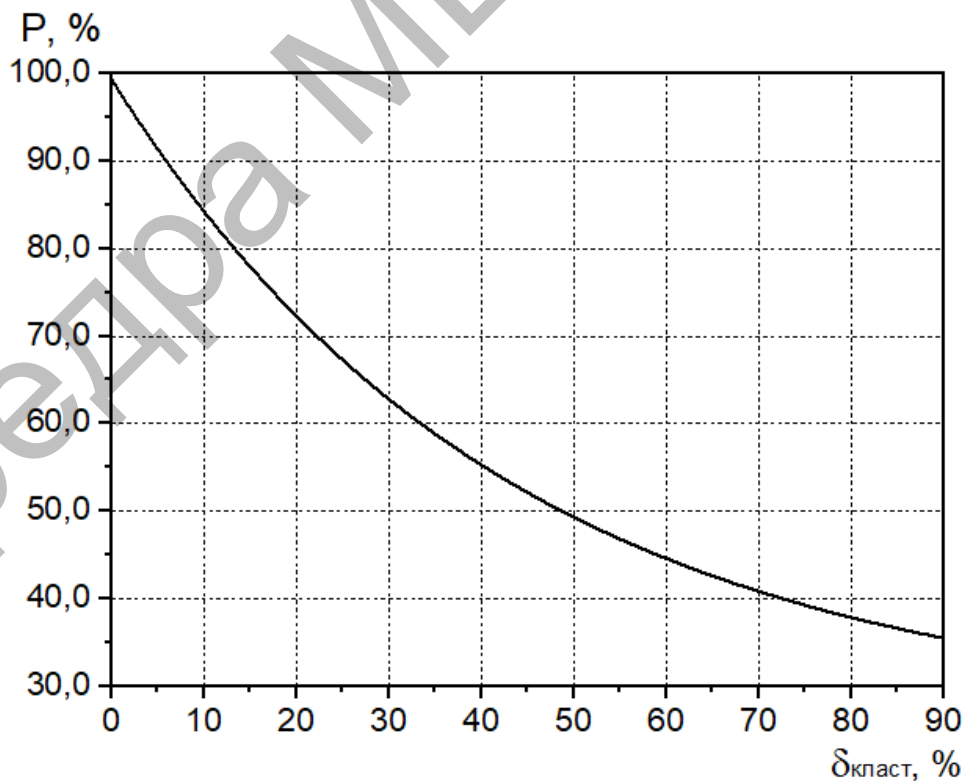


Рисунок 3.11 – Залежність достовірності прийняття рішення від похибок кластеризації

Результати роботи програми приведені на рис. 3.11. Так, при зміні відносної похибки кластеризації досліджуваних масивів вимірюного тиску від 0 % до 90 % ймовірність прийняття правильного рішення відповідно змінюється в межах від 100 % до 36 %. Слід відмітити, що в процесі досліджень використовувались дані вимірювань тиску одного і того ж технологічного процесу. Тому добитися 100 % похибок кластеризації не вдалося.

3.5 Висновки до розділу

В результаті досліджень впливу похибок кластеризації масивів даних вимірювань на достовірність прийняття рішень в трьох каналній інтелектуальній вимірювальній інформаційній системі тиску встановлено, що:

— для оцінювання впливу похибок кластеризації на достовірність прийняття рішення можна використати метрики якості кластеризації та подальші аналізи. Однією з таких метрик є Adjusted Rand Index (ARI).

4 ОХОРОНА ПРАЦІ І НАВКОЛИШНЬОГО СЕРЕДОВИЩА

Правовою основою законів про охорону праці є: Конституція України; Закон України "Про охорону праці"; Кодекс законів про працю України. і ін. [19].

Основним законодавчим документом у сфері охорони праці є Закон України "Про охорону праці", дія якого поширюється на всі підприємства, установи та організації, незалежно від форм власності та видів діяльності, а також на всіх громадян, які працюють і зайняті на цих підприємствах. Він складається з 44 статі, з'єднаної в 9 розділів.

Темою даної дипломної роботи є дослідження похибок кластеризації фреймів даних вимірювань в багатопараметричних просторово розподілених інтелектуальних вимірювальних інформаційних системах. Всі дослідження проводяться за допомогою комп'ютера.

Як відомо, робота з комп'ютером, не забезпечена певними організаційно-технічними захисними заходами, безумовно, негативно позначається на здоров'я і самопочутті користувачів. У зв'язку з цим актуальними є питання захисту користувачів персональних електронно-обчислювальних машин ПЕОМ від шкідливих дій, що виникають в процесі роботи з комп'ютерною технікою.

Особи, які професійно використовують комп'ютери, повинні проходити обов'язковий попередній медичний огляд (під час прийняття на роботу) і періодичні медичні огляди в організованому порядку та у встановлені терміни. До безпосередньої роботи з комп'ютером допускаються особи, які не мають медичних протипоказань. Вагітні та жінки, які годують груддю, не можуть виконувати будь-які види робіт на комп'ютері.

Професійні користувачі комп'ютерів можуть працювати, якщо їм виповнилося 18 років, вони мають право працювати з електрообладнанням і пройшли всі види інструктажу.

У робочих зонах, де працюють з комп'ютерами, повинні бути аптечки першої допомоги та вуглекислотні вогнегасники.

4.1 Виробнича санітарія

Санітарні питання регулюються українським законодавством, Законом "Про забезпечення санітарного та епідемічного благополуччя населення", який встановлює права та обов'язки державних органів, підприємств, установ, організацій та громадян у сфері забезпечення санітарного та епідемічного благополуччя населення, встановлений порядок організації роботи державної санітарно-епідеміологічної служби, здійснення санітарного нагляду в Україні

Виробничі приміщення повинні проектуватися відповідно до вимог ДБН В.2.2-28:2010 "Будинки і споруди". та СН 512-78 "Вказівки з проектування будинків і приміщень для обчислювальних машин".

Площа одного робочого місця повинна становити щонайменше 6,0 квадратних метрів, а об'єм – щонайменше 20 кубічних метрів, з урахуванням максимальної кількості людей, які одночасно працюють у зміні.

Комп'ютерні зали не повинні розташовуватися під землею. На внутрішніх входах і виходах не повинно бути порогів. Якщо рівні підлоги в сусідніх приміщеннях відрізняються, то в місцях їх переміщення повинні бути передбачені похилі поверхні. Поверхня підлоги в комп'ютерних приміщеннях повинна бути рівною, без отворів, неслизькою, легко митися, невологою та антистатичною.

Для внутрішнього оздоблення приміщень дифузні світловідбиваючі матеріали з коефіцієнтом відбиття стелі від 0,7 до 0,8, стін від 0,5 до 0,6 і підлоги від 0,3 до 0,5 повинні бути дозволені до використання органами Державного санітарно-епідеміологічного нагляду.

Планування робочого місця, де встановлені ПК, повинно враховувати мінімальну відстань 2,0 м між робочими столами з моніторами (у напрямку від задньої панелі одного монітора до екрану іншого монітора) та мінімальну відстань 1,2 м між боковими сторонами моніторів.

При проектуванні робочих місць слід враховувати їх кількість і конструктивні особливості (розміри робочих столів і комп'ютерів, клавіатур, подіумів тощо), характер виконуваних робіт і забезпечувати оптимальне

розташування використовуваного обладнання на робочій поверхні. При цьому можуть використовуватися робочі столи різних конструкцій, що відповідають сучасним ергономічним вимогам.

Конструкція робочих стільців (крісел) повинна забезпечувати підтримку розумної робочої пози під час роботи за комп'ютером і дозволяти користувачеві змінювати позу для зняття статичної напруги в м'язах шиї, плечей і спини та запобігання втомі.

Тип робочого стільця (крісла) слід вибирати відповідно до характеру і тривалості комп'ютерної роботи з урахуванням зросту користувача. Робочий стілець (крісло) повинен мати можливість регулювання висоти і кута нахилу сидіння і спинки, а також відстані спинки від переднього краю сидіння шляхом підйому, опускання і обертання, причому регулювання кожного параметра повинно бути незалежним, простим у виконанні і надійно фіксованим.

Приміщення з ПК повинні бути обладнані системами опалення та кондиціонування повітря або ефективною припливно-витяжною вентиляцією; вимоги до вентиляції, опалення та кондиціонування повітря в приміщеннях з ПК виконуються відповідно до будівельних норм.

Повітря, яке поступає в приміщення ПЕОМ, слід очищати від забруднення, у тому числі від пилу і мікроорганізмів .

Параметри мікроклімату повинні бути наступними :

– холодна пора року: температура повітря від 22 °С до 24 °С; відносна вологість від 60 % до 40 %;

– влітку: температура повітря від 21 °С до 25 °С; відносна вологість від 60 % до 40 %.

У приміщеннях з персональними комп'ютерами ПК використовуйте зволожувачі повітря для підвищення вологості повітря, що заправляються щодня дистильованою або прокип'яченою питною водою.

Допустимі рівні звукового тиску, шуму та еквівалентного рівня шуму на робочих місцях повинні відповідати гігієнічним нормам виробничого шуму, ультразвуку та інфразвуку.

Для зниження рівня шуму і вібрації в комп'ютерних приміщеннях обладнання та пристрої повинні встановлюватися на спеціальних фундаментах або віброгасильних прокладках, як це передбачено нормативними документами. Понизити рівень шуму в приміщеннях з ПЕОМ можна також використанням звукопоглинальних матеріалів з максимальними коефіцієнтами звукопоглинання в області частот 63,8 Гц для оздоблення приміщень, перевірених спеціальними акустичними розрахунками. Щоб збільшити звукопоглинання, повісьте щільні однотонні штори, які гармоніюють з кольором стін, склавши їх на відстані від 15 см до 20 см від паркану. Ширина штори повинна бути вдвічі більшою за ширину вікна. Якщо основна комп'ютерна робота виконується в приміщенні, рівень шуму на робочому місці не повинен перевищувати 50 дБ (А). Коли інженери і технічний персонал працюють в лабораторіях, аналітичних кабінетах і кімнатах управління вимірюваннями, рівень шуму не повинен перевищувати 60 дБ (А).

Вібрація обладнання на робочому місці не повинна перевищувати допустимих значень, встановлених державними санітарними нормами.

Освітлення в приміщеннях з ПЕОМ повинно бути поєднанням природного і штучного світла; робочі зони з ВДТ і ПЕОМ повинні розташовуватися по відношенню до прорізів так, щоб природне світло падало збоку, бажано зліва. Штучне освітлення в приміщеннях з ПЕОМ повинно забезпечуватися загальною системою рівномірного освітлення. У виробничих, адміністративних і громадських приміщеннях, де виконується великий обсяг роботи з документами, допускається застосування систем комбінованого освітлення (загальне освітлення плюс додаткові місцеві світильники, призначені для освітлення зони розташування документів). Освітленість поверхонь столів у місцях розташування робочих документів повинна становити від 300 до 500 лк. Для освітлення документів дозволяється використовувати світильники місцевого освітлення. Місцеве освітлення не повинно створювати відблисків на поверхні екрана та підвищувати освітленість екрана більше ніж на 300 люкс. Прямі відблиски від джерел освітлення повинні бути обмежені, а яскравість

світловипромінюючих поверхонь (наприклад, вікон, світильників) у полі зору не повинна перевищувати 200 кд/кв.м. Відбиті відблиски на робочих поверхнях (екранах, столах, клавіатурах тощо) слід обмежувати шляхом вибору відповідного типу світильників і розташуванням робочої зони по відношенню до джерел природного і штучного освітлення, яскравість відблисків на екрані комп'ютера не повинна перевищувати 40 кд/кв.м, а відбите освітлення – 200 кд/кв.м. Яскравість стелі не повинна перевищувати 200 кд/кв.м під час використання системи. Індекс засліпленості для загального штучного освітлення у виробничих приміщеннях повинен бути менше 20, індекс дискомфорту для адміністративних і громадських приміщень – менше 40, індекс дискомфорту для дитячих садків і навчальних закладів – менше 25. Нерівномірність розподілу яскравості в полі зору користувача комп'ютера повинна бути обмежена, при цьому співвідношення яскравості між робочими поверхнями не повинно перевищувати 3:1 і 5:1, а співвідношення між робочими поверхнями і поверхнями стін і обладнання – 10:1. При штучному освітленні в якості джерела світла повинні використовуватися переважно люмінесцентні лампи типу ЛБ. Для відбивного освітлення у виробничих приміщеннях та адміністративно-громадських будівлях можуть використовуватися металогалогенні лампи потужністю до 250 Вт. Лампи розжарювання можуть використовуватися для місцевого освітлення. Загальне освітлення слід передбачати у вигляді суцільних або переривчастих ліній світильників з боків робочих місць, паралельно лінії зору користувача, якщо комп'ютери розташовані в ряд. Якщо комп'ютери розташовані по периметру, лінію світильників слід розташовувати ближче до переднього краю робочого столу, зверненого до оператора.

Для освітлення приміщення з ПК слід використовувати світильники з дзеркальними розсіювачами серії ЛПО36, оснащені високочастотними пускорегулювальними апаратами (ВЧ-пускорегулювальні апарати). Використання світильників без розсіювачів та екрануючих решіток заборонено. Яскравість світильників загального освітлення в зонах з кутом випромінювання

від 50° до 90° до вертикальної і горизонтальної площини не повинна перевищувати 200 кд/кв.м, а захисний кут світильників повинен становити не менше 40° . Світильники для місцевого освітлення повинні бути обладнані відбивачами, що не відбивають світло, з кутом захисту не менше 40° . Коефіцієнт запасу (K_z) для світильників загального освітлення повинен становити 1,4. Коефіцієнт пульсації не повинен перевищувати 5 %. Це повинно забезпечуватися використанням газорозрядних ламп для світильників загального освітлення та місцевих світильників з високочастотними пускорегулювальними апаратами (ВЧ-пускорегулювальними апаратами) для всіх типів світильників. За відсутності світильників з ВЧ-баластами лампи в багатолампових світильниках або сусідніх світильниках загального призначення слід вмикати на різні фази трифазної мережі.

Для забезпечення нормальної освітленості в приміщеннях з комп'ютерами віконні рами та скло ламп слід очищати не рідше двох разів на рік, а перегорілі лампи – своєчасно замінювати.

У приміщеннях з комп'ютерами слід використовувати нейтралізатори для запобігання та захисту від статичної електрики. Захист від статичної електрики повинен бути забезпечений відповідно до гігієнічних норм щодо допустимої напруженості електричних полів. Допустимий рівень напруженості електростатичного поля не повинен перевищувати 20 Вт на годину.

Візуальна дисплейна техніка генерує декілька видів випромінювання, включаючи рентгенівське, радіочастотне та ультрафіолетове випромінювання, але рівні цих випромінювань є досить низькими і не перевищують норми.

Робочий час і час відпочинку при роботі з комп'ютером повинен бути організований відповідно до виду і категорії трудової діяльності. Тривалість обідньої перерви визначається чинним законодавством про працю та правилами внутрішнього трудового розпорядку підприємства (організації, установи). З метою забезпечення оптимальної працездатності та збереження здоров'я професійних користувачів повинні надаватися встановлені періоди відпочинку

протягом робочої зміни. Тривалість перерв для відпочинку протягом робочої зміни встановлюється залежно від її тривалості, виду та категорії трудової діяльності. Під час регламентованих перерв слід виконувати комплекс спеціальних вправ для зниження нервово-емоційного напруження, втоми зорового аналізатора, усунення наслідків гіпонатріємії та гіпонатріємії, запобігання настанню постнапруженої втоми. Для зменшення негативних наслідків моніторингу доцільно чергувати змістовні маніпуляції з текстом і числовими даними (зміна змісту роботи) та редагування тексту і введення даних (зміна змісту роботи).

Якщо, незважаючи на дотримання гігієнічних та ергономічних вимог і режиму праці та відпочинку, особа, яка працює з комп'ютером, відчуває зоровий дискомфорт або інші небажані суб'єктивні відчуття, слід обмежити час роботи з комп'ютером, скоригувати періоди відпочинку або слід застосовувати індивідуальні підходи, наприклад, переключитися на інші види діяльності, не пов'язані з використанням комп'ютера.

4.2 Безпека праці

Перед початком роботи. При вимкненому вимикачі живлення і роз'єми живлення видаліть пил з екрана і поверхні монітора, принтера і процесорного блоку за допомогою злегка вологої ганчірки. Виконайте візуальний огляд пристрою, шнура живлення та інтерфейсу. При виявленні механічних пошкоджень на корпусі, екрані монітора, шнури живлення або інтерфейсі користувач повинен негайно звернутися до сервісного центру. Підключивши штекер живлення до електромережі, увімкніть пристрій за допомогою відповідного вимикача на корпусі. Не підключайте і не від'єднуйте роз'єм живлення від електромережі, якщо вимикач на корпусі пристрою увімкнений. Якщо пристрій було вимкнено, увімкніть його знову протягом однієї хвилини. Підключайте та відключайте інтерфейсний кабель між пристроями тільки після того, як обидва пристрої будуть вимкнені.

Під час роботи принтера слідкуйте за станом чорнильної стрічки та якістю друку.

Вимикайте монітори, комп'ютерні блоки живлення, принтери та інше офісне обладнання, а також від'єднуйте роз'єми живлення після завершення роботи.

Не експлуатуйте обладнання:

- Обладнання, яке було визнано експертом несправним або непридатним для використання;
- Обладнання, яке не підключено до контуру заземлення;
- Будь-які механічні пошкодження корпусу або дисководу;
- Будь-який виріб з пошкодженими пломбами;
- Несправний блок живлення.

4.3 Пожежна безпека

Правовою основою діяльності у сфері пожежної безпеки є Конституція, Закон України "Про пожежну безпеку", інші закони України, постанови Верховної Ради України, укази і розпорядження Президента України, декрети, постанови і розпорядження Кабінету Міністрів України, рішення органів державної виконавчої влади, рішення, прийняті в межах компетенції органів місцевого та автономного самоврядування. При забезпеченні пожежної безпеки слід також керуватися українськими правилами пожежної безпеки, стандартами, будівельними нормами, правилами улаштування електроустановок (ПУЕ), нормами технологічного проектування та іншими нормативно-правовими актами, що регламентують вимоги пожежної безпеки відповідно до сфери їх застосування.

Закон України "Про пожежну безпеку", прийнятий 17 грудня 1993 року, є основним нормативним документом, що регулює вимоги пожежної безпеки. Закон визначає загальні правові, економічні та соціальні засади забезпечення

пожежної безпеки в Україні і регулює відносини між державними органами, юридичними та фізичними особами у цій сфері незалежно від виду діяльності та форми власності.

Відповідно до ДБН В.1.1-7:2016 "Пожежна безпека в будівництві" та ДСТУ-Н Б В.1-7:2016 "Загальні вимоги пожежної безпеки в будівництві. Загальні вимоги" та ДСТУ-Н Б А.3.2-1:2007 "Система стандартів безпеки праці. Настанова з визначення небезпечних і шкідливих факторів під час виробництва будівельних матеріалів і виробів та їх застосування під час будівництва і експлуатації об'єктів будівництва та захисту від їх впливу", ймовірність виникнення пожежі або вибуху протягом одного року не повинна перевищувати від 10 до 6.

Будівлі та частини будівель, де встановлюються ПК, повинні мати ступінь вогнестійкості не нижче II. Усі будівлі та приміщення, в яких експлуатуються відеотермінали та ПК, повинні мати класифікацію за вибухопожежною та пожежною небезпекою згідно з ДСТУ Б В.1.1-36:2016 "Визначення категорії приміщень, будинків та зовнішніх установок за вибухопожежною та пожежною небезпекою", а також класифікацію зон відповідно до правил улаштування електроустановок повинні бути присвоєні. Відповідне позначення повинно бути нанесене на вхідні двері приміщення. Основними причинами виникнення та поширення пожеж є недотримання правил використання та експлуатації обладнання та приладів зі слабким протипожежним захистом, застосування окремих випадкових матеріалів, що не відповідають протипожежним вимогам, а також відсутність ефективних систем пожежогасіння.

При гасінні пожежі застосовують рідкі, паро-, газо-, піно-образні і тверді речовини. Паро- і газоподібні речовини, вуглекислота, водяна пара й інертні гази, поширюючи і забираючи на своє нагрівання тепло гарячого середовища і швидко змішуючись з гарячими газами, витісняють кисень із зони горіння, створюють середовище, у якому неможливо не тільки горіння, але й тління. Інертні гази при гасінні не псують продуктів і устаткування. Піноутворюючі речовини знайшли саме широке застосування при гасінні не тільки твердих

матеріалів, але і палих і легкозаймистих рідин. Піна покриває поверхню палаючих матеріалів, ізолює від кисню, охолоджує і припиняє горіння. Піноутворюючі речовини отримуються у результаті хімічної реакції або процесу механічного перемішування, і являють собою пухирці газу (вуглекислоти) чи повітря, укладені в тонкі водяні механічні міцні плівки.

Тверді речовини, вуглекислота (вуглекислий сніг), виконують охолоджуючу дію, ізолюють палаючу поверхню від кисню і припиняють горіння. Вуглекислий сніг застосовують для гасіння майже всіх хімічних речовин. При з'єднанні з палаючими речовинами він не утворює шкідливих з'єднань. Для комп'ютерної техніки застосування рідких і пінних вогнегасників обмежено, треба використовувати ручні вуглекислотні вогнегасники, наприклад, ОУ-2.

Приміщення, де розміщуються персональні комп'ютери та дисплейні зали, повинні бути обладнані автоматичною пожежною сигналізацією з димовими пожежними сповіщувачами та переносними вуглекислотними вогнегасниками з розрахунку два на 20 м^2 площі підлоги з урахуванням гранично допустимої концентрації вогнегасної речовини. Не рідше одного разу на квартал слід витирати пил з вузлів і агрегатів, кабельних каналів і міжповерхових перекриттів.

Пожежна профілактика включає в себе заходи, спрямовані на запобігання пожежам, блокування шляхів розповсюдження вогню та забезпечення швидкої і безпечної евакуації людей і майна у разі виникнення пожежі. Ці заходи включають розгортання протипожежного обладнання та чітку організацію протипожежних заходів пожежною командою.

4.4 Розрахунок заземлюючого пристрою

Опір природного заземлення складає $R_e = 10 \text{ Ом}$, допустимий опір заземлюючого пристрою $R_0 = 4 \text{ Ом}$, питомий опір ґрунту $\rho = 87 \text{ Ом}\cdot\text{м}$. Коефіцієнт сезонності $\psi = 1,3$. Глибина залягання електроду $h = 0,5 \text{ м}$.

Контур заземлення виконують із сталевих стрижнів, кутів, некондиційних труб і ін. В траншеї глибиною до 0,7 м вертикально забиваються стрижні (труби, кути і ін.), а виступаючі із землі верхні кінці з'єднуються зваркою поверх сталевую смугою або прутком.

При цьому необхідно дотримуватися наступних умов:

- поперечний переріз з'єднувальної планки повинен бути не менше 48 мм² і не менше 4 мм завтовшки.;
- найменший діаметр проводу – 10 мм
- довжина стрижня повинна бути не менше 1,5 м, щоб досягти незамерзаючого шару ґрунту.

Тому як вертикальні електроди приймаємо прутки діаметром $d = 0,02$ м і завдовжки $l = 3,5$ м.

Знаходимо допустимий опір штучного заземлюючого пристрою, Ом

$$R_{ш} = \frac{R_e \cdot R_o}{R_e - R_o} = \frac{10 \cdot 4}{10 - 4} = 6,67$$

Знаходимо відстань від поверхні землі до середини вертикального електроду, м

$$t = h + \frac{l}{2} = 0,5 + 1,75 = 2,25$$

Приймаємо відстань між вертикальними електродами $a = 3,5$ м.

Знаходимо опір одиночного вертикального заземлюючого пристрою, Ом

$$R_e = \frac{\rho \cdot \psi}{2 \cdot \pi \cdot l_e} \cdot \left(\ln \frac{2 \cdot l_e}{d} + \frac{1}{2} \cdot \ln \frac{4 \cdot t + l_e}{4 \cdot t - l_e} \right)$$

$$R_e = \frac{87 \cdot 1,3}{2 \cdot 3,14 \cdot 3,5} \cdot \left(\ln \frac{2 \cdot 3,5}{0,02} + \frac{1}{2} \cdot \ln \frac{4 \cdot 2,25 + 3,5}{4 \cdot 2,25 - 3,5} \right) = 31$$

Знаходимо орієнтовне число вертикальних заземлюючих пристроїв

$$n^{орієнт} = \frac{R}{R_e} = \frac{31}{6,67} = 4,6$$

Знаходимо орієнтовний коефіцієнт використання вертикальних електродів

$$\eta_e^{\text{опієт}} = 0,7$$

Знаходимо число вертикальних заземлюючих пристроїв

$$n = \frac{R_e}{R_{\text{ш}} \cdot \eta_e} = \frac{31}{6,67 \cdot 0,7} = 6,63$$

Округляємо число електродів до 7 і знаходимо коефіцієнт використання вертикальних електродів

$$\eta_e = 0,73$$

Знаходимо довжину горизонтального електроду. При розташуванні електродів в ряд довжина горизонтального електроду, м

$$l_z = a \cdot (n - 1) = 3,5 \cdot (7 - 1) = 21$$

Приймаємо товщину горизонтального електроду $b = 0,005$ м.

Знаходимо опір горизонтального електроду

$$R_z = \frac{\rho \cdot \psi}{2 \cdot \pi \cdot l_z} \cdot \ln \frac{2 \cdot l_z^2}{b \cdot h}; \quad R_z = \frac{87 \cdot 1,3}{2 \cdot 3,14 \cdot 21} \cdot \ln \frac{2 \cdot 21^2}{0,005 \cdot 0,5} = 11,0$$

Знаходимо коефіцієнт використання горизонтального електроду

$$\eta_z = 0,65$$

Знаходимо опір штучного заземлюючих пристроїв заземлюючого пристрою, Ом

$$R_{\text{ш}} = \frac{R_e \cdot R_z}{R_e \cdot \eta_z + n_e \cdot R_z \cdot \eta_e} = \frac{31 \cdot 11}{31 \cdot 0,73 + 7 \cdot 11 \cdot 0,65} = 4,69$$

Знаходимо загальний опір заземлюючого пристрою, Ом

$$R = \frac{R_e \cdot R_{\text{ш}}}{R_e + R_{\text{ш}}} = \frac{10 \cdot 4,69}{10 + 4,69} = 3,19$$

Оскільки опір заземлюючого пристрою менш $R_0 = 4$ Ом розрахунок виконаний, вірно.

Проаналізувавши результати даного дипломного проекту був виявлений негативний вплив на людину, а також проведений розрахунок заземлюючого пристрою, використання якого зменшить вплив на робітників такого небезпечного фактору, як ураження електричним струмом.

Запропоновані організаційно-технічні заходи при роботі з персональними комп'ютерами дозволили поліпшити умови праці і зменшити травматизм на розглянутих робочих місцях.

4.5 Висновок до розділу

Запропоновані організаційно-технічні заходи при роботі з персональними комп'ютерами дозволили поліпшити умови праці і зменшити травматизм на розглянутих робочих місцях.

Кафедра МБЖД ХНУДУ

ВИСНОВКИ

В результаті виконання дипломної роботи була вирішена науково-практична задача з дослідження впливу похибок кластеризації фреймів даних вимірювань на достовірність прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних системах.

Проведені дослідження дозволили зробити наступні висновки:

1. Для успішного аналізу результатів кластеризації важливо мати необхідні й достатні умови:

— необхідно використовувати відповідні метрики для оцінки якості кластеризації;

— необхідно візуалізувати дані та кластери для отримання загального уявлення про те, наскільки добре вони розділені;

— необхідно перевіряти, чи правильно вибрані параметри для алгоритму кластеризації (наприклад, кількість кластерів), використання неправильних параметрів може призвести до неправильних результатів;

— необхідно перевіряти стабільність результатів кластеризації, використовуючи різні підвибірки або запуски алгоритму з різними початковими умовами;

— завжди потрібно перевіряти, що отримані кластери можуть бути використані для вирішення задачі класифікації даних вимірювань у просторово розподілених вимірювальних інформаційних системах;

Загалом, тільки комбінація кількох аспектів дозволить здійснити комплексний аналіз результатів кластеризації і визначити їхню ефективність.

2. Похибки кластеризації можуть виникати внаслідок розподілених за різними законами розподілу даних. Деякі з причин цих похибок включають: неспільність розподілів, наявність шуму, масштабні різниці, наявність «важких хвостів», викиди. Внаслідок наявності цих дестабілізуючих факторів нормована похибка кластеризації масивів даних на виході досліджуваної просторово

розподіленої інтелектуальної вимірювальної інформаційної системи лежать в межах від 0,4 до 0,7.

3. Оцінка впливу похибок кластеризації даних вимірювань на достовірність прийняття рішень в інтелектуальних вимірювальних інформаційних системах може бути складною задачею, оскільки вона включає в себе якість самої кластеризації, так і вплив цих похибок на вирішення конкретних завдань системи. Встановлено, при зміні відносної похибки кластеризації досліджуваних масивів вимірюного тиску від 0 % до 90 % ймовірність прийняття правильного рішення в просторово розподіленій інтелектуальній вимірювальній інформаційній системі тиску відповідно змінюється в межах від 100 % до 36 %.

Таким чином, в результаті проведених досліджень мета роботи досягнута, а завдання на магістерську роботу виконано у повному обсязі.

Запропоновані в даній дипломній роботі рішення доцільно використовувати в процесі наукових досліджень просторово розподілених інтелектуальних вимірювальних інформаційних систем та в навчальному процесі.

ПЕРЕЛІК ПОСИЛАНЬ

1. Wohwe Sambo, D., Yenke, B. O., Förster, A., & Dayang, P. (2019). Optimized clustering algorithms for large wireless sensor networks: A review. *Sensors*, 19(2), 322..
2. Parizad, A., & Hatziaioniu, C. J. (2022). Cyber-attack detection using principal component analysis and noisy clustering algorithms: A collaborative machine learning-based framework. *IEEE Transactions on Smart Grid*, 13(6), 4848-4861.
3. Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178-210.
4. L. J. P. van der Maaten and G. E. Hinton. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* 9(Nov):2579-2605, 2008.
5. G. E. Hinton and S. T. Roweis. Stochastic Neighbor Embedding. In *Advances in Neural Information Processing Systems*, volume 15, pages 833–840, Cambridge, MA, USA, 2002. The MIT Press.
6. R. A. Jacobs. Increased rates of convergence through learning rate adaptation. *Neural Networks*, 1: 295–307, 1988.
7. L. J. P. van der Maaten. Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research* 15(Oct):3221-3245, 2014.
8. Visualizing t-SNE Clustering. URL:<https://www.naftaliharris.com/blog/visualizing-t-sne-clustering>.
9. 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>.
10. 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>.

11. 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>.
12. 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>.
13. Sheppard, Kevin. 2023. "Linearmodels: Instrumental Variable and Linear Panel Models for Python." <https://pypi.org/project/linearmodels/>.
14. 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>.
15. Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. The MIT Press. <http://www.jstor.org/stable/j.ctt5hhcfr>.
16. Parizad, A., & Hatziadoniu, C. J. (2022). Cyber-attack detection using principal component analysis and noisy clustering algorithms: A collaborative machine learning-based framework. *IEEE Transactions on Smart Grid*, 13(6), 4848-4861.
17. Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178-210.
18. Коваль А. О., Коваль О. А. Просторово розподілені інтелектуальні вимірювальні інформаційні системи: монографія. Харків : Лідер, 2017. 146 с.
19. Закон України "Про охорону праці" в редакції від 21 листопада 2002 р. URL: https://urst.com.ua/act/pro_okhoronu_pratsi (дата звернення: 16.09.2023).

ДОДАТОК А
ІЛЮСТРАТИВНИЙ МАТЕРІАЛ ДО ДИПЛОМНОЇ РОБОТИ

Кафедра МЕЖДУНАРОДУ

Харківський національний автомобільно-дорожній університет
Факультет механічний
Кафедра метрології та безпеки життєдіяльності

ІЛЮСТРАТИВНИЙ МАТЕРІАЛ ДО ДИПЛОМНОЇ РОБОТИ
магістра

ДОСЛІДЖЕННЯ ВПЛИВУ ПОХИБОК КЛАСТЕРИЗАЦІЇ ФРЕЙМІВ ДАНИХ ВИМІРЮВАНЬ НА ДОСТОВІРНІСТЬ ПРИЙНЯТТЯ РІШЕНЬ В БАГАТОПАРАМЕТРИЧНИХ ПРОСТОРОВО РОЗПОДІЛЕНИХ ІНТЕЛЕКТУАЛЬНИХ ВИМІРЮВАЛЬНИХ СИСТЕМАХ

Завідувач кафедри, канд. техн. наук, доц.

Нормоконтролер, канд. техн. наук, доц.

Керівник, канд. техн. наук, доц.

Студент гр. ММ-61-22


О. І. Богатов


І. В. Грайворонська


О. А. Козаль


Д. В. Грабовський

м. Харків - 2023

МЕТА, ОБ'ЄКТ ТА ПРЕДМЕТ ДОСЛІДЖЕНЬ

Об'єкт дослідження — просторово розподілена трьохканальна інтелектуальна вимірювальна інформаційна система тиску.

Мета роботи — дослідити вплив похибок кластеризації фреймів даних вимірювань на достовірність прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних системах.

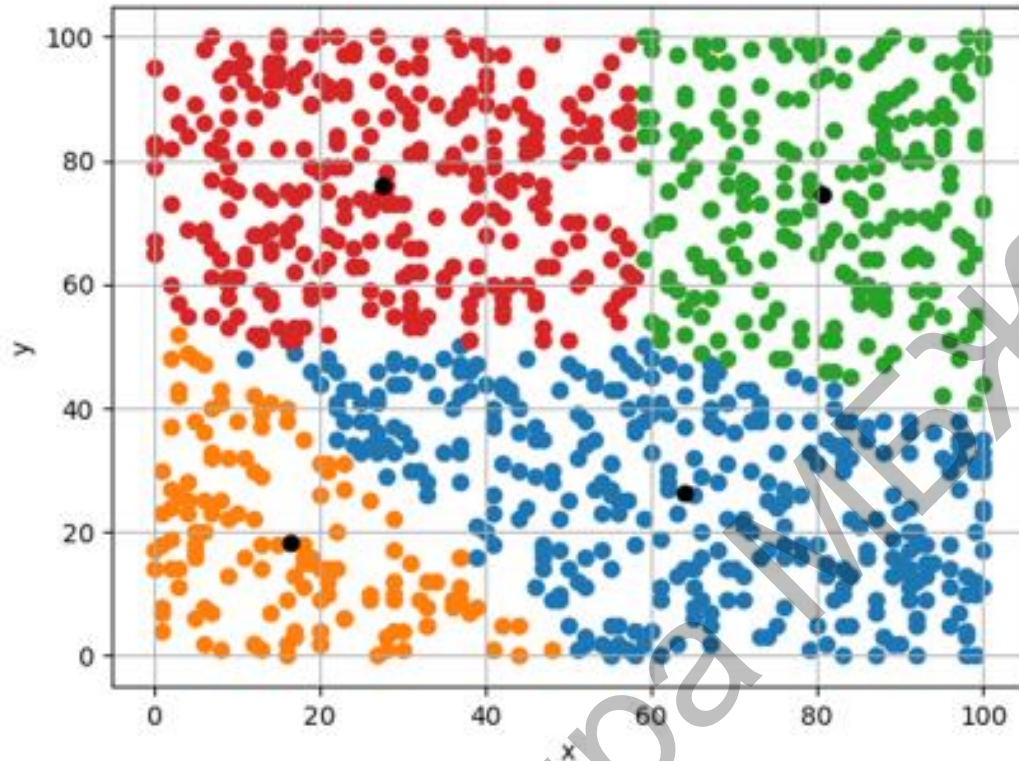
Метод дослідження — кластерного, статистичного та часового аналізу даних вимірювань просторово розподіленої інтелектуальної вимірювальної інформаційної системи тиску.

Предмет дослідження — процес кластеризації вимірюваних масивів даних тиску.

В магістерській роботі проведені дослідження впливу похибок кластеризації фреймів даних вимірювань на достовірність прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних системах.

Кафедра МБХІНДУ

РЕАЛІЗАЦІЯ КЛАСТЕРИЗАЦІЇ К-СЕРЕДНІХ



$$p = \sqrt{\sum_{i=1}^n (x_n - x_{n \text{ centr}})^2} \quad (1)$$

Рисунок 1 – Кластеризація для точок двовимірного простору.

РЕАЛІЗАЦІЯ КЛАСТЕРИЗАЦІЇ К-СЕРЕДНІХ

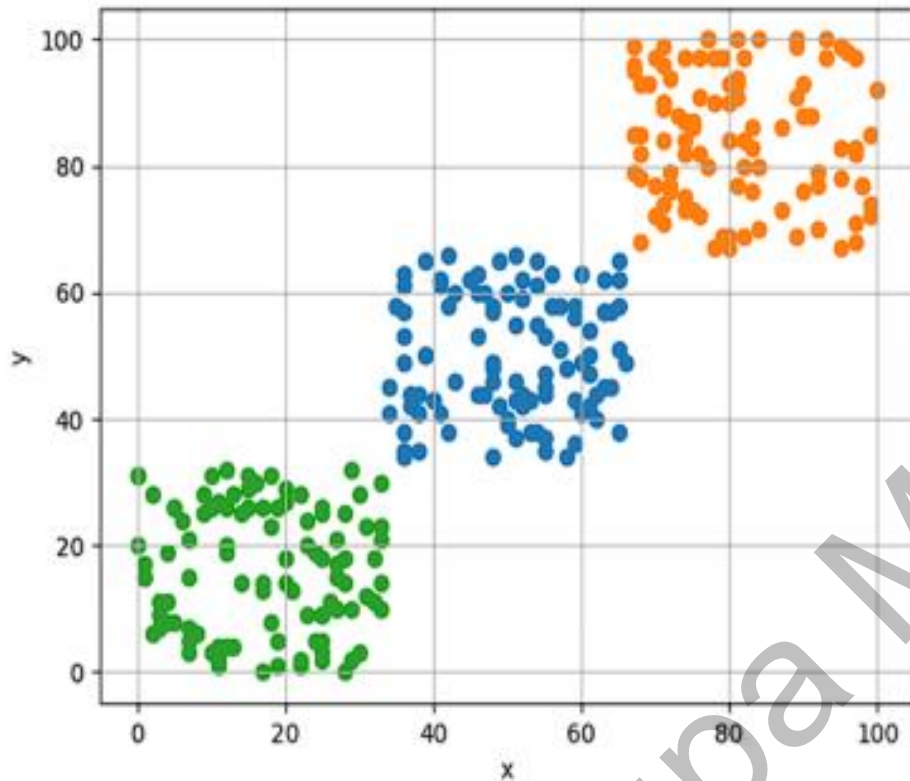


Рисунок 2 – Кластеризація для точок двовимірного простору. Кількість точок – 300. Кількість кластерів – 3. Дані згруповано.

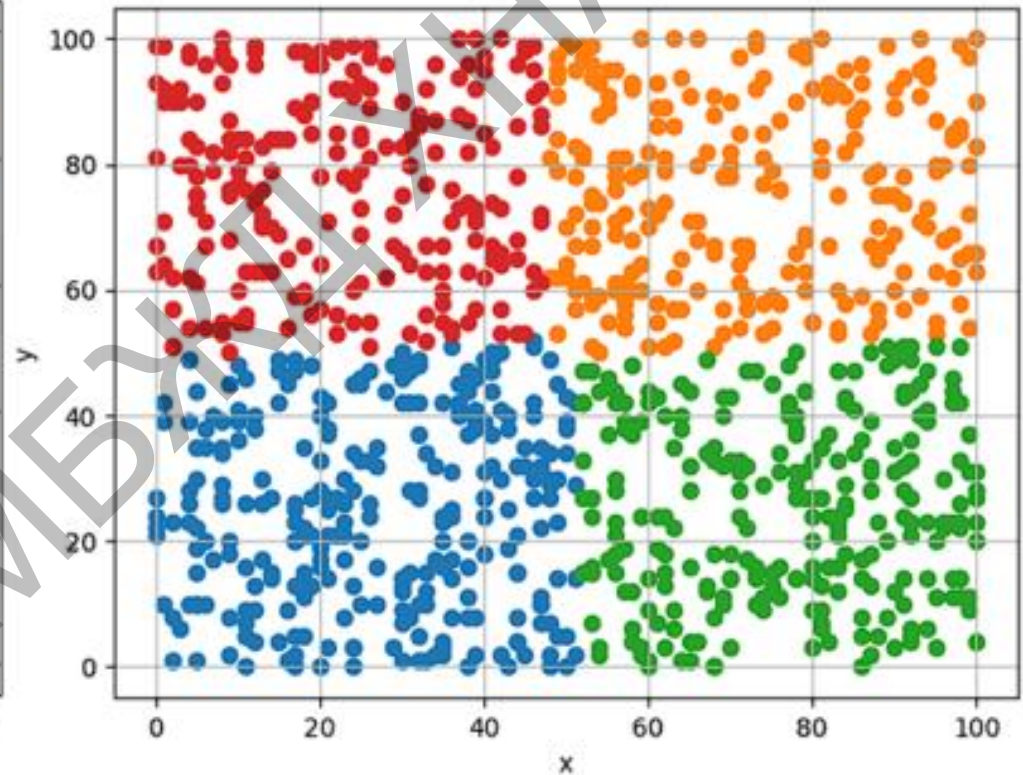


Рисунок 3 – Кластеризація точок двовимірного простору. Кількість точок – 1000. Кількість кластерів – 4. Дані не згруповані.

РЕАЛІЗАЦІЯ КЛАСТЕРИЗАЦІЇ К-СЕРЕДНІХ

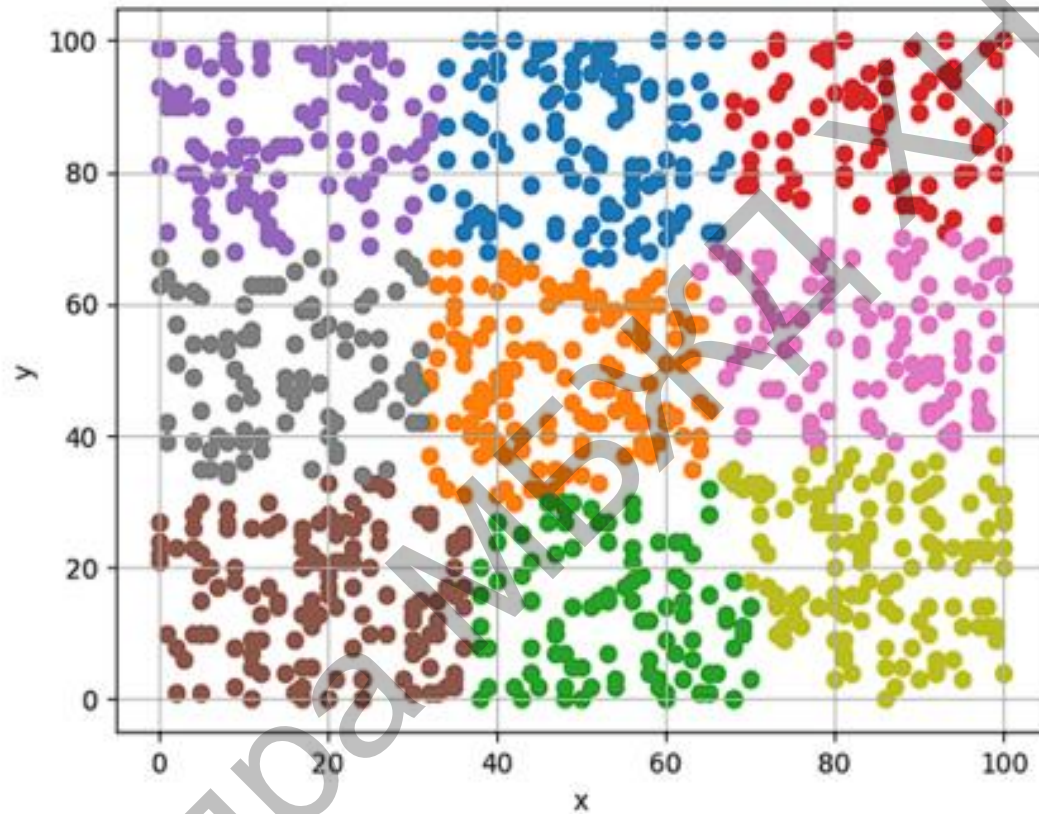


Рисунок 4 – Кластеризація точок двовимірного простору.
Кількість точок 1000. Кількість кластерів – 9. Дані не згруповані.

РЕАЛІЗАЦІЯ КЛАСТЕРИЗАЦІЇ К-СЕРЕДНІХ

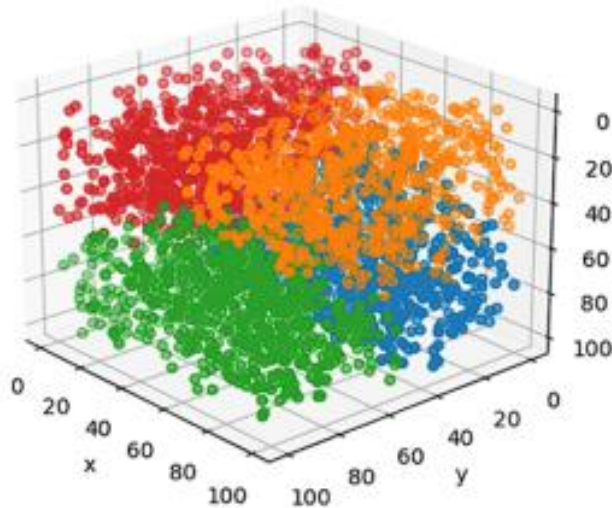


Рисунок 5 – Кластеризації для 3-х мірного простору. Кількість точок – 3000. Кількість кластерів – 4. Дані не згруповані.

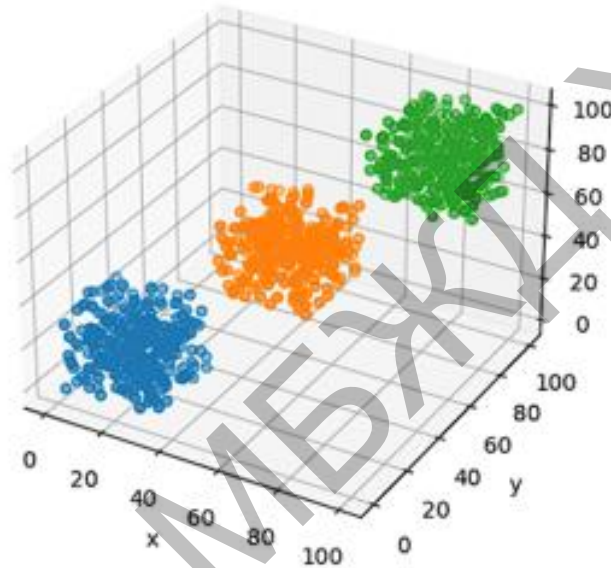


Рисунок 6 – Кластеризації тривимірного простору для згрупованих даних. Кількість точок – 900. Кількість кластерів – 3

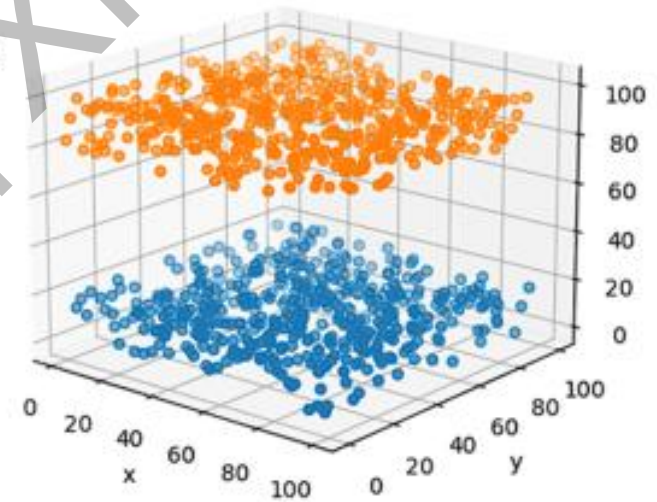


Рисунок 7 – Кластеризація тривимірного простору. Кількість точок – 1000. Кількість кластерів – 2. Дані згруповано.

Кафедра

ДАТЧИК ТИСКУ САПФІР-22ДТ-ВН (2440)



Перетворювач вимірювальний різниці тисків Сапфір-22ДТ-Вн, модель 2440 з верхньою межею вимірювань 250 кПа і гранично допустимим робочим надлишковим тиском 16 МПа. Мембранний, з мембранами зі сплаву 36НХТЮ, з фланцями з вуглецевої сталі, що має вид кліматичного виконання УХЛЗ.1, для роботи при температурі навколишнього повітря від 5 до 50°С з межею допустимої основної похибки $\gamma = \pm 0,25\%$, з вихідним сигналом 0-5 мА, з ніпелями, варіант встановлення перетворювача - на плиті, з вентиляним блоком.

Рисунок 8 – Аналоговий датчик тиску Сапфір-22ДТ-Вн (2440)

РЕЗУЛЬТАТИ КЛАСТЕРИЗАЦІЇ МЕТОДОМ К-СЕРЕДНІХ ВИМІРЯНОГО ТИСКУ ІВІС З ДАТЧИКОМ САПФІР-22ДТ-ВН (2440)

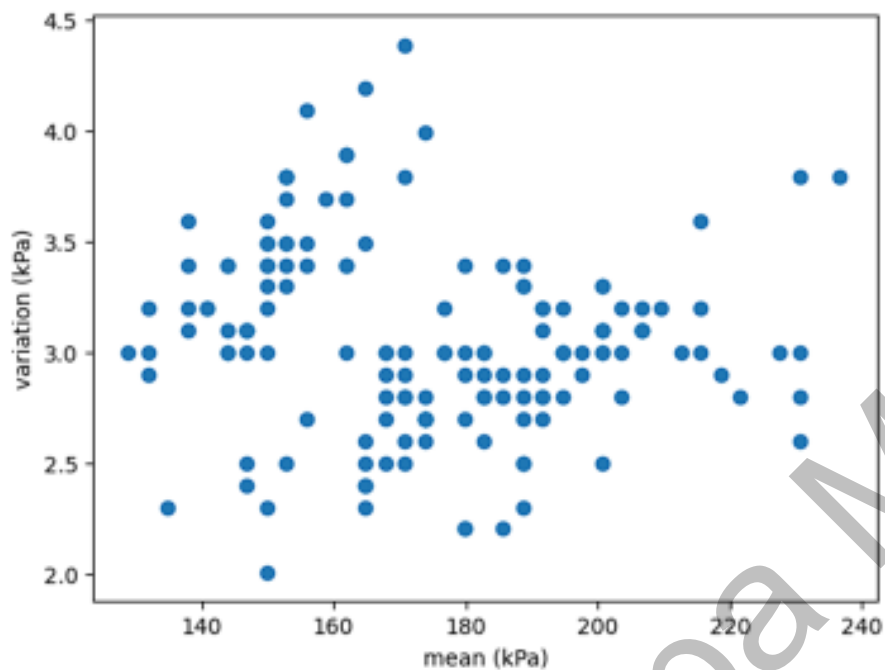


Рисунок 9 – Необроблені вхідні дані

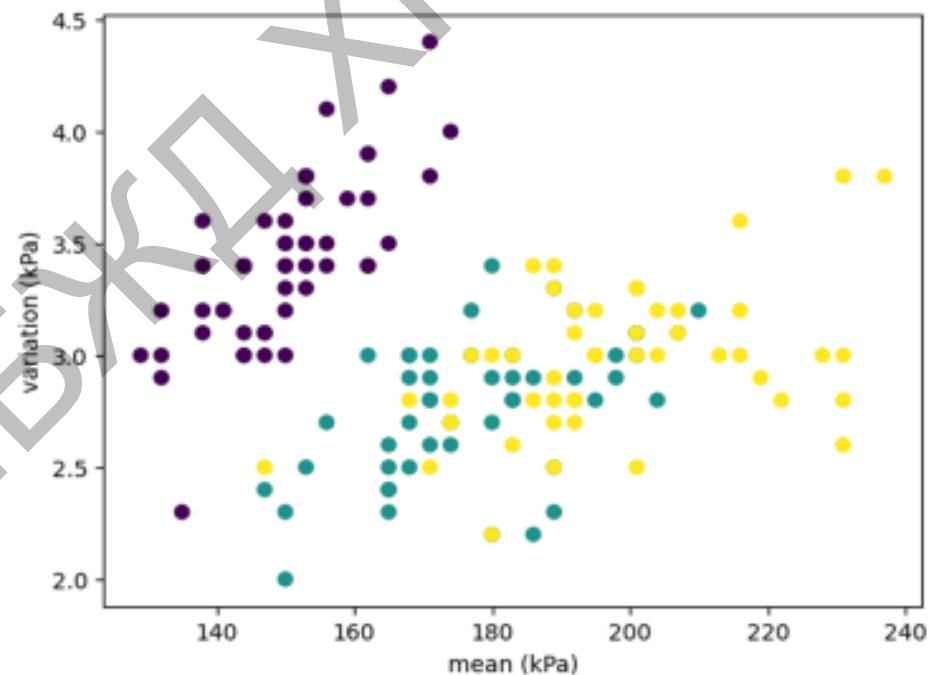


Рисунок 10– Результати кластеризації виміряного тиску датчиків за середнім значенням та дисперсією виміряного тиску

РЕЗУЛЬТАТИ КЛАСТЕРИЗАЦІЇ ВИМІРЯНОГО ТИСКУ ІВІС З ДАТЧИКОМ САПФІР-22ДТ-ВН (2440)

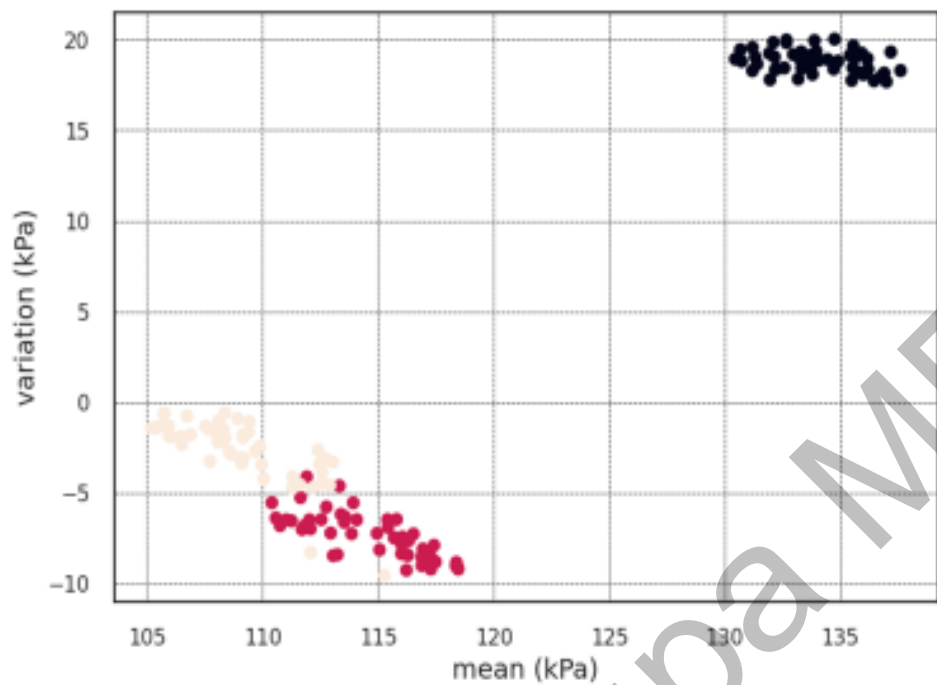


Рисунок 11 – Результат кластеризації методом t-SNE

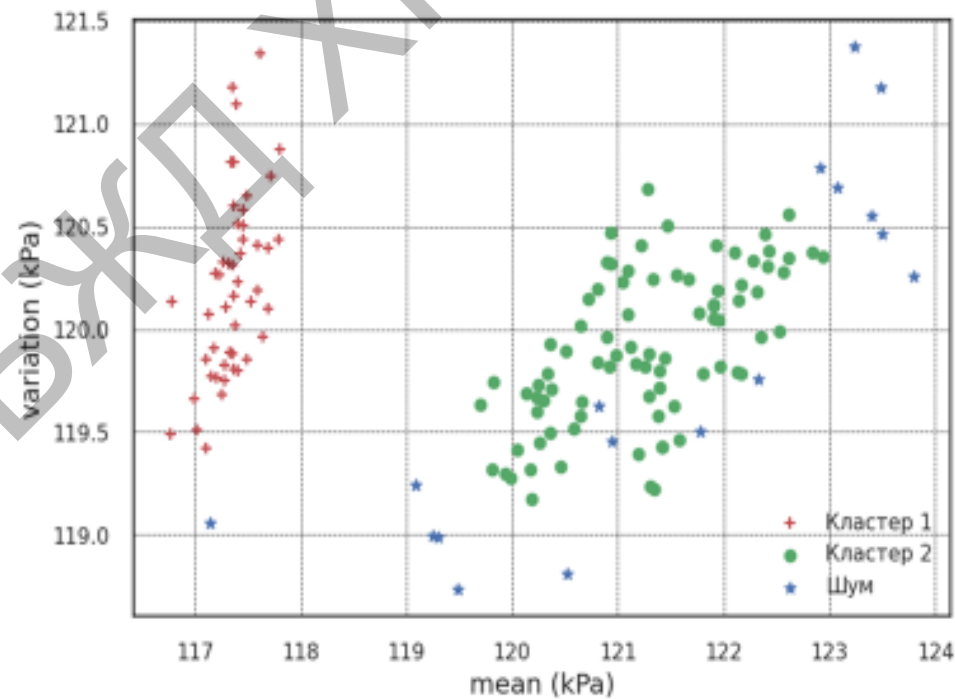


Рисунок 12 – Результати роботи алгоритму DBSCAN

Кафедра МБЖД ХНУДУ

РЕЗУЛЬТАТИ КЛАСТЕРИЗАЦІЇ ВИМІРЯНОГО ТИСКУ ІВІС З ДАТЧИКОМ САПФІР-22ДТ-ВН

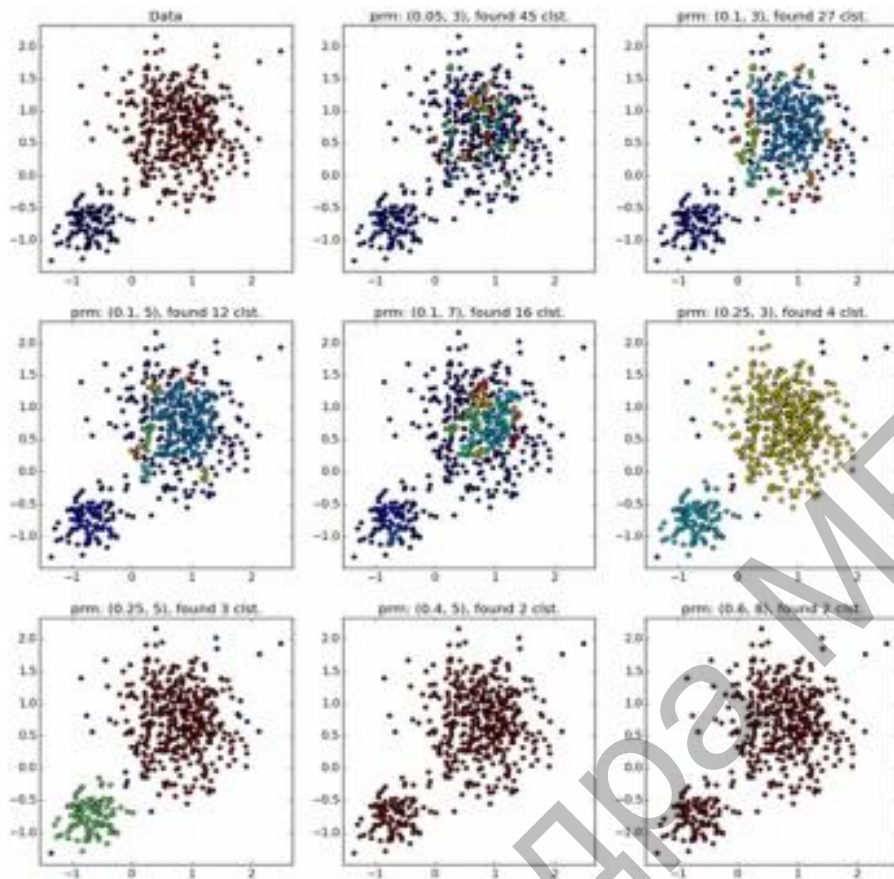


Рисунок 12 – Кластеризація скупчення даних вимірювань тиску різного розміру

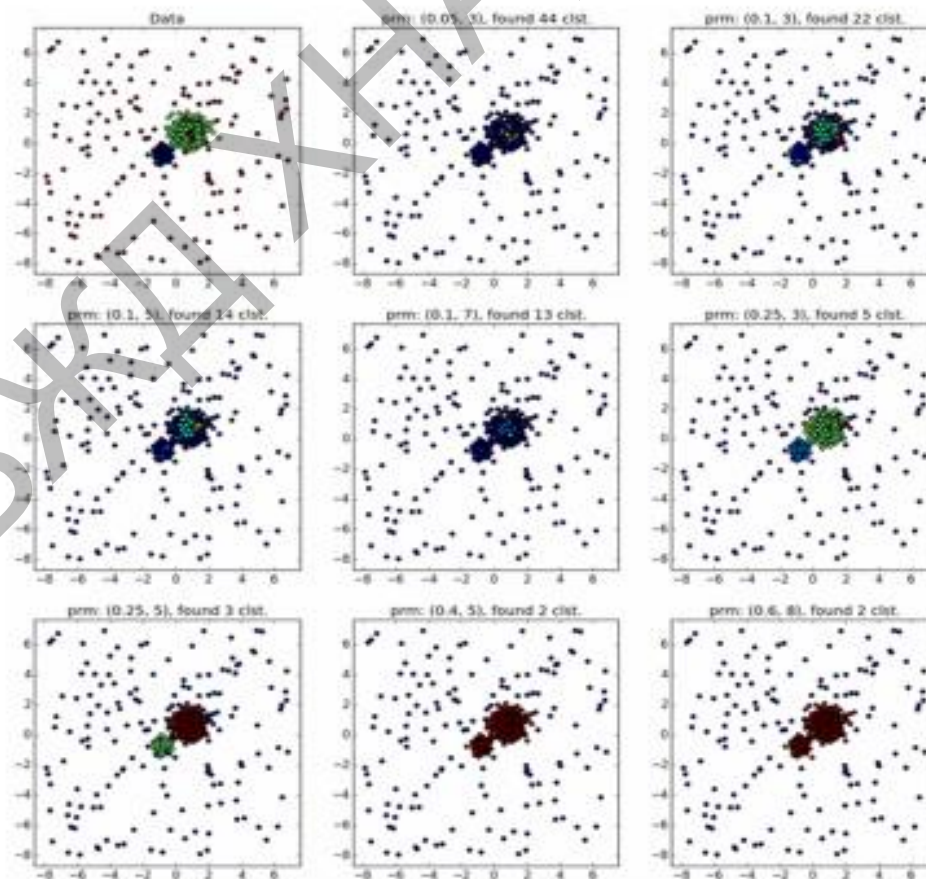


Рисунок 13 – Кластеризація сильно зашумлених даних

РЕЗУЛЬТАТИ КЛАСТЕРИЗАЦІЇ ВИМІРЯНОГО ТИСКУ ІВІС З ДАТЧИКОМ САПФІР-22ДТ-ВН

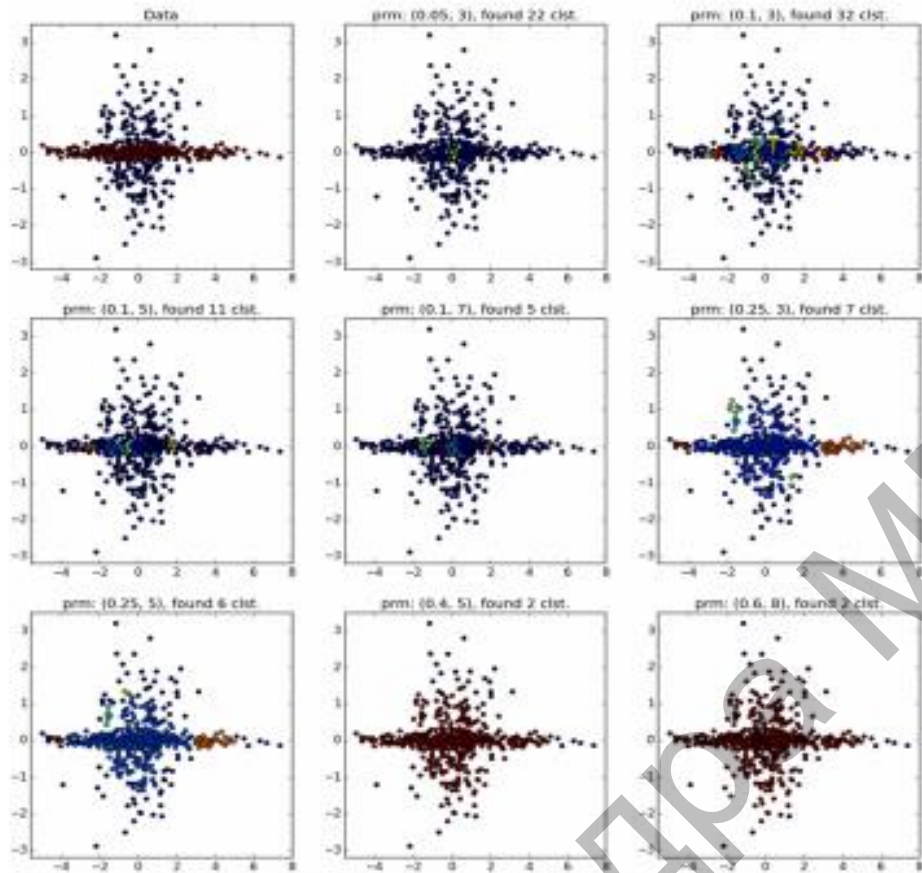


Рисунок 14 – Кластеризація скупчення даних вимірювань тиску різної щільності

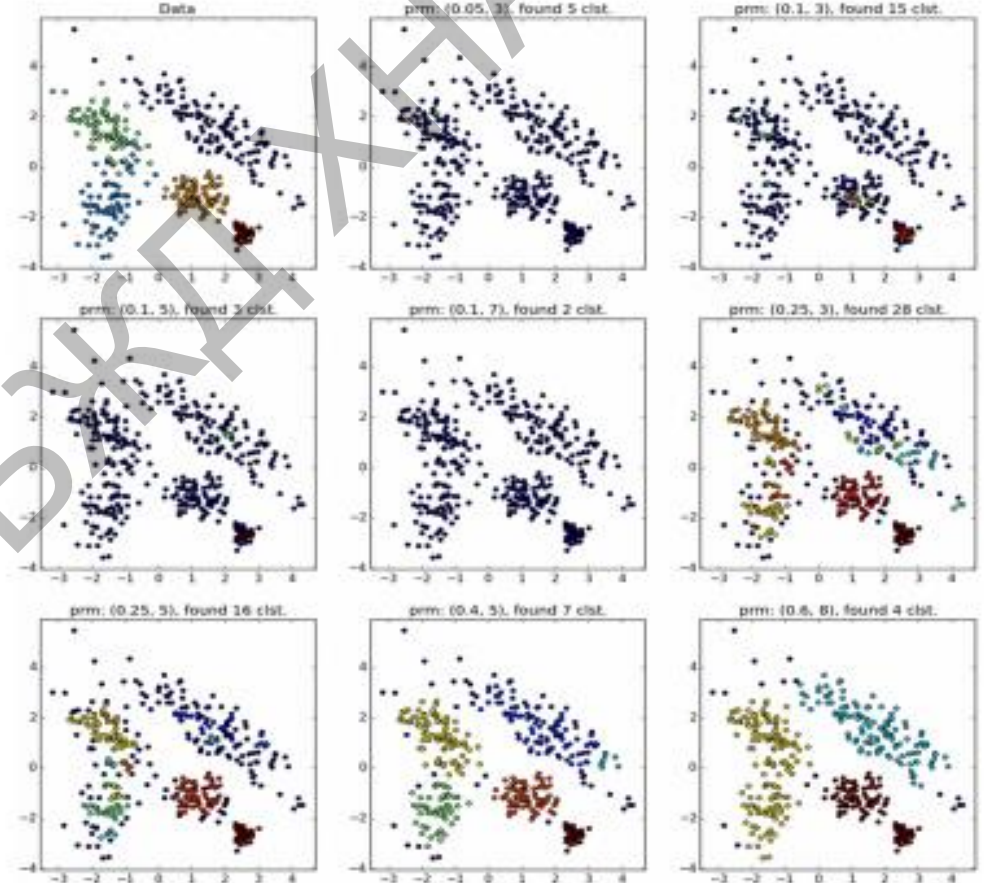


Рисунок 15 – Кластеризація масивів даних тиску з різними розподілами

КАФЕДРА МЕЖДІСЦИПЛІНАРНОЇ НАУКИ

ОСОБЛИВОСТІ ВИМІРЯНОГО ТИСКУ ІВІС З ДАТЧИКОМ САПФІР-22ДТ-ВН

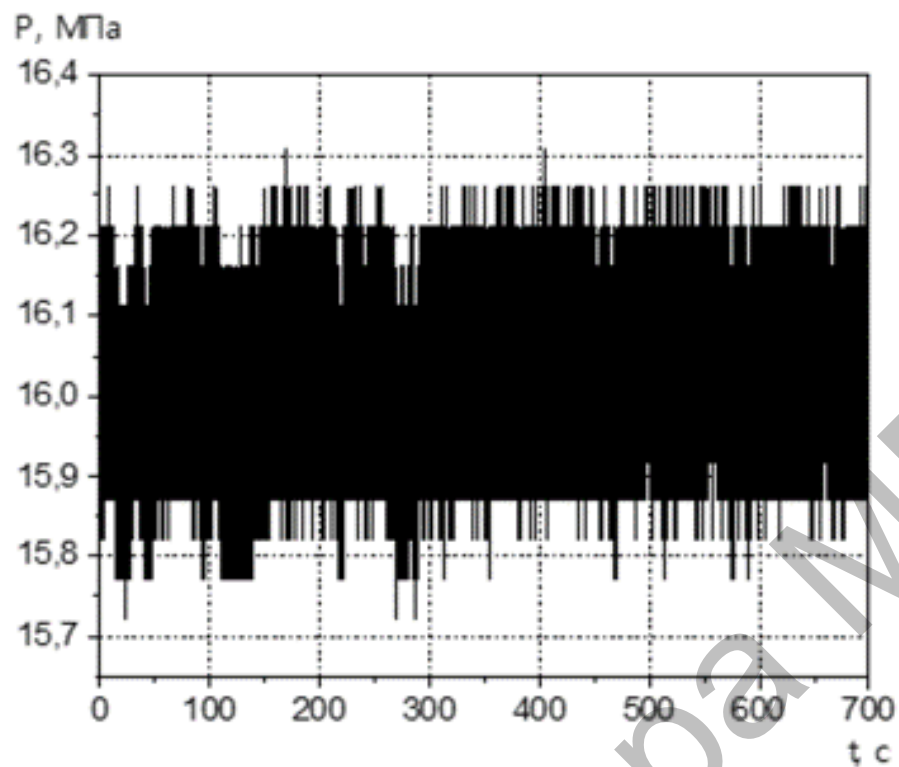


Рисунок 16 – Характер виміряного тиску в нічний час доби

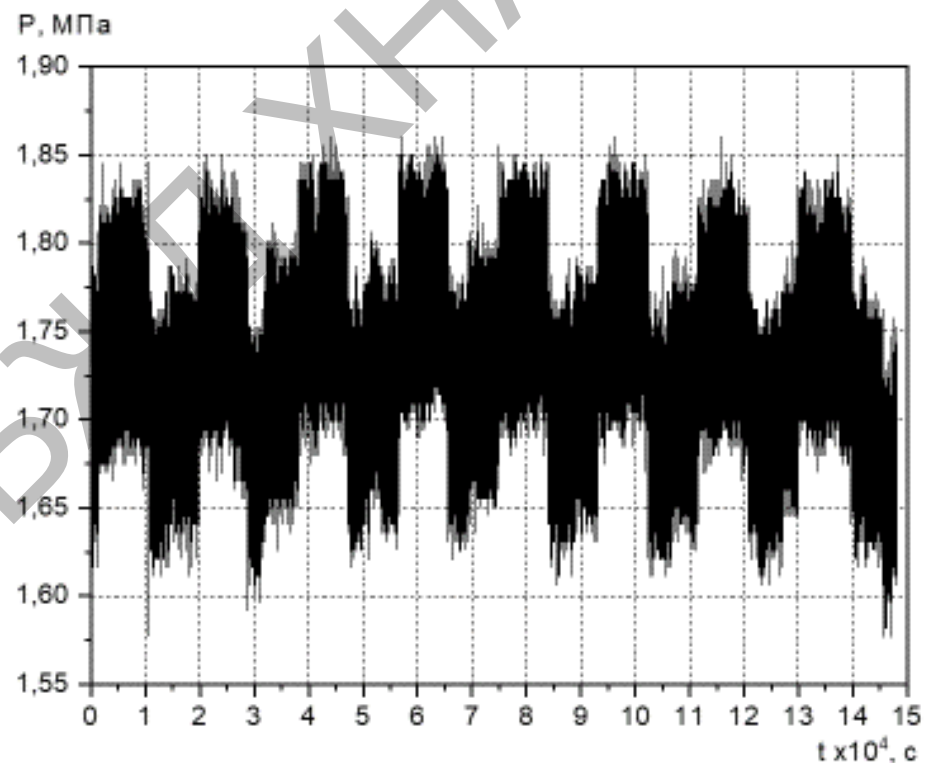


Рисунок 17 – Характер виміряного тиску в період з 06.00 до 22.00

МАТРИЦІ ВІДСТАНЕЙ КЛАСТЕРИЗАЦІЇ ДАНИХ ВИМІРЮВАНЬ ІВІС ТИСКУ

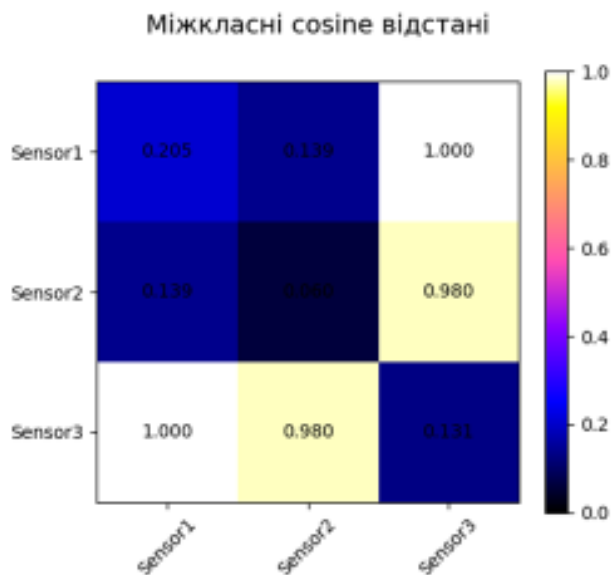


Рисунок 18 – Міжкласні косинусні відстані

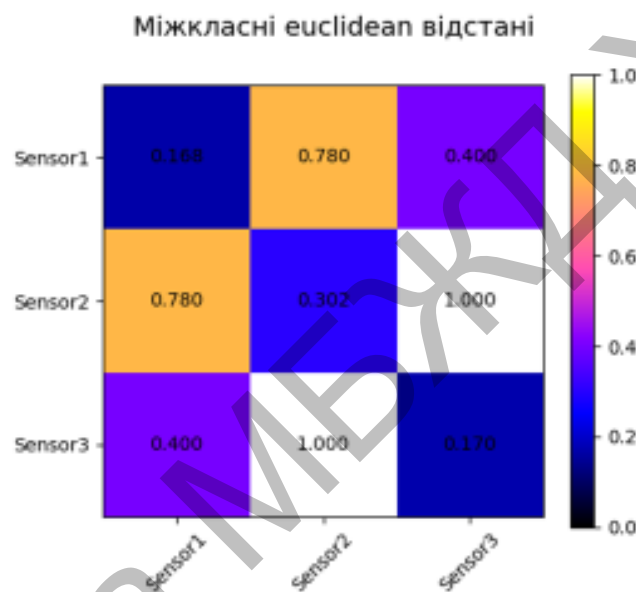


Рисунок 19 – Міжкласні евклідові відстані

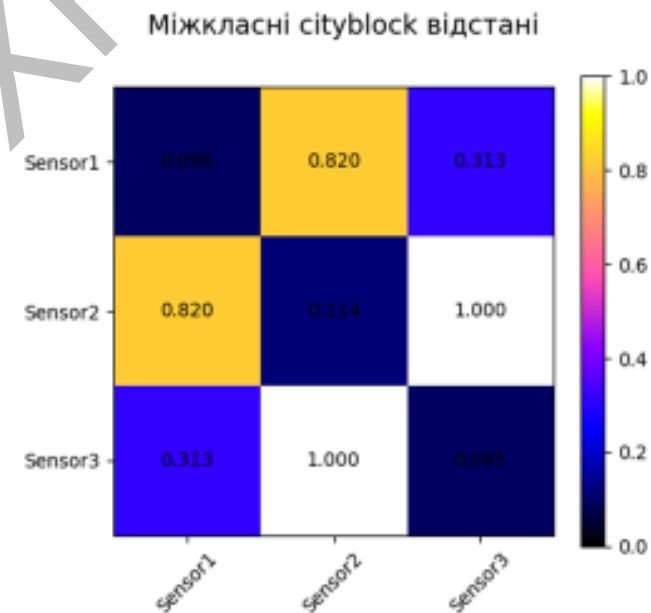


Рисунок 20 – Міжкласні cityblock відстані

РЕЗУЛЬТАТИ АНГЛОМЕРАТИВНОЇ КЛАСТЕРИЗАЦІЇ ВИМІРЯНОГО ТИСКУ ІВІС З ДАТЧИКОМ САПФІР-22ДТ-ВН

Агломеративна кластеризація (спорідненість=cosine)

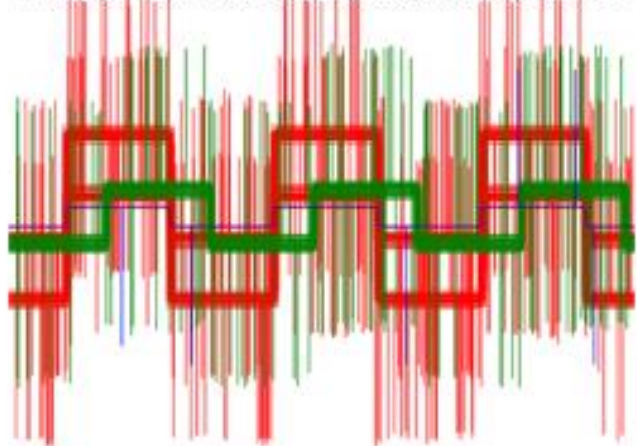


Рисунок 21 – Англомеративна кластеризація (cosine спорідненість)

Агломеративна кластеризація (спорідненість=euclidean)

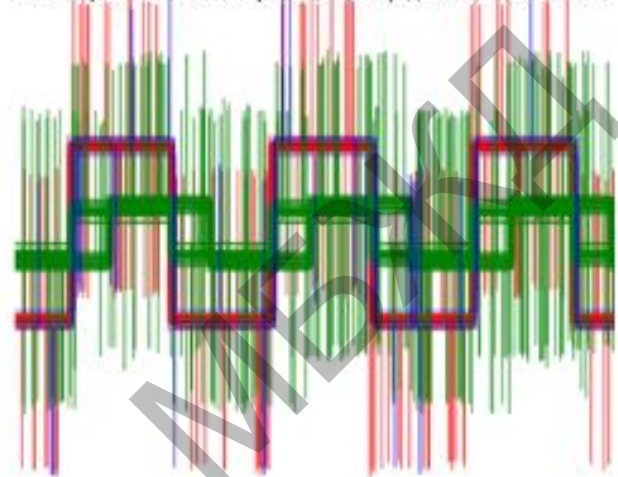


Рисунок 22 – Англомеративна кластеризація (euclidean спорідненість)

Агломеративна кластеризація (спорідненість=cityblock)

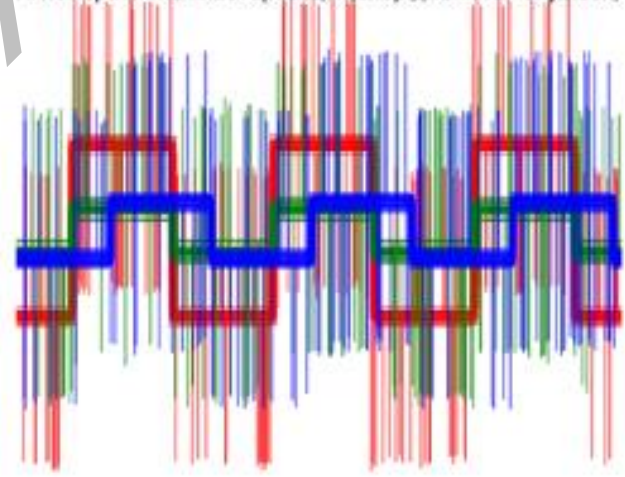


Рисунок 23 – Англомеративна кластеризація (cityblock спорідненість)

Кафедра МБМД ХНУДУ

РЕЗУЛЬТАТИ АНГЛОМЕРАТИВНОЇ КЛАСТЕРИЗАЦІЇ ВИМІРЯНОГО ТИСКУ ІВІС З ДАТЧИКОМ САПФІР-22ДТ-ВН

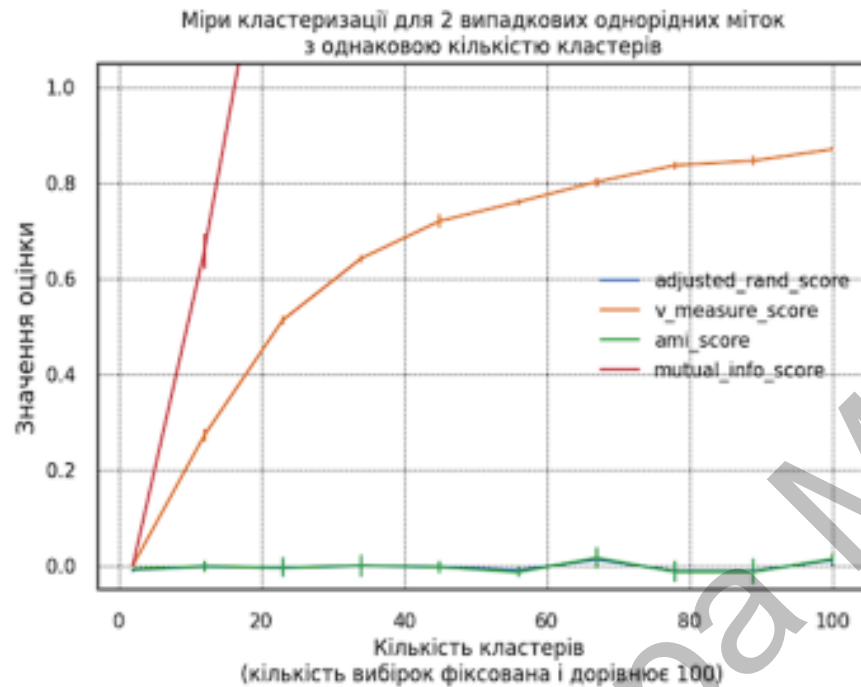


Рисунок 24 – Міри кластеризації для двох однорідних масивів виміряного тиску з однаковою кількістю кластерів

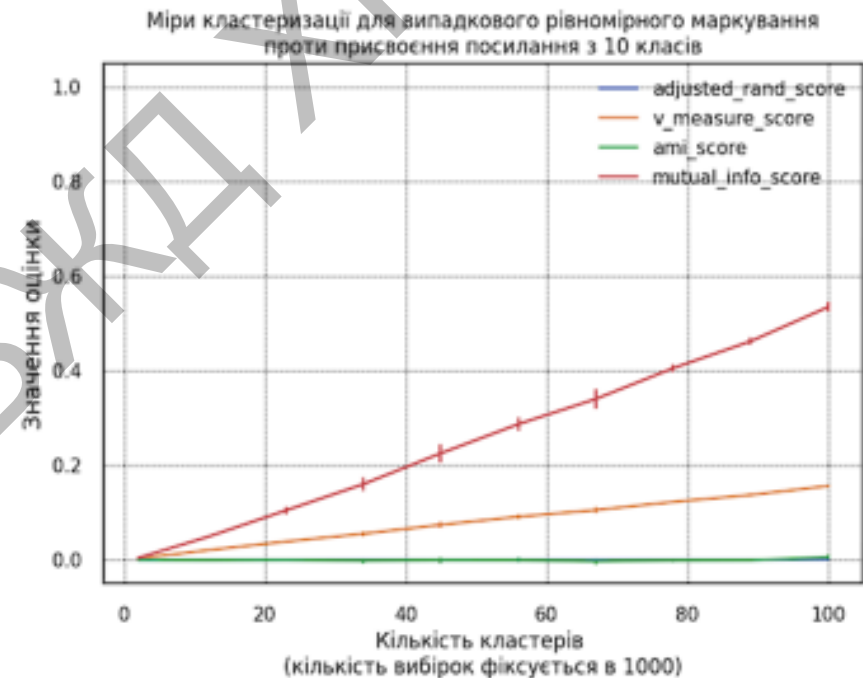


Рисунок 25 – Міри кластеризації для двох випадкових масивів виміряного тиску з однаковою кількістю кластерів

ЗАЛЕЖНІСТЬ ДОСТОВІРНОСТІ ПРИЙНЯТТЯ РІШЕННЯ ВІД ПОХИБОК КЛАСТЕРИЗАЦІЇ І ВІС ТИСКУ

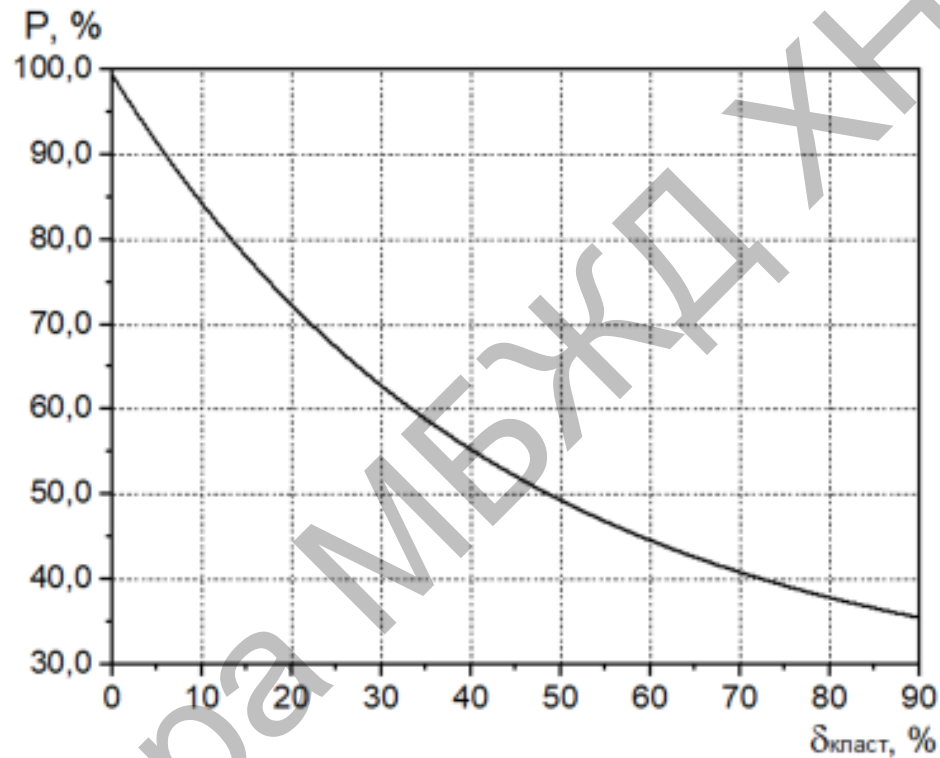


Рисунок 26 – Залежність достовірності прийняття рішення від похибок кластеризації

Кафедра МБЖД ХНУДУ

ВИСНОВКИ

В результаті виконання дипломної роботи була вирішена науково-практична задача з дослідження впливу похибок кластеризації фреймів даних вимірювань на достовірність прийняття рішень в багатопараметричних просторово розподілених інтелектуальних вимірювальних системах.

Проведені дослідження дозволили зробити наступні висновки:

1. Похибки кластеризації можуть виникати внаслідок розподілених за різними законами розподілу даних. Деякі з причин цих похибок включають: неспільність розподілів, наявність шуму, масштабні різниці, наявність «важких хвостів», викиди. Внаслідок наявності цих дестабілізуючих факторів нормована похибка кластеризації масивів даних на виході досліджуваної просторово розподіленої інтелектуальної вимірювальної інформаційної системи лежать в межах від 0,4 до 0,7.

2. Оцінка впливу похибок кластеризації даних вимірювань на достовірність прийняття рішень в інтелектуальних вимірювальних інформаційних системах може бути складною задачею, оскільки вона включає в себе якість самої кластеризації, так і вплив цих похибок на вирішення конкретних завдань системи. Встановлено, при зміні відносної похибки кластеризації досліджуваних масивів виміряного тиску від 0% до 90% ймовірність прийняття правильного рішення в просторово розподіленій інтелектуальній вимірювальній інформаційній системі тиску відповідно змінюється в межах від 100% до 36%.

Таким чином, в результаті проведених досліджень мета роботи досягнута, а завдання на магістерську роботу виконано у повному обсязі.

Запропоновані в даній дипломній роботі рішення доцільно використовувати в процесі наукових досліджень просторово розподілених інтелектуальних вимірювальних інформаційних систем та в навчальному процесі.