

ГЕНТОШ ЛЕСЯ

Національний університет «Львівська політехніка»

<https://orcid.org/0000-0002-4957-1512>e-mail: lesia.i.mochurad@lpnu.ua**ЛЕВКОВИЧ РОМАН**

Національний університет «Львівська політехніка»

<https://orcid.org/0000-0001-9393-714X>e-mail: roman.y.levkovych@lpnu.ua

ПАРАЛЕЛЬНИЙ ПІДХІД ДО КЛАСТЕРИЗАЦІЇ БУДІВЕЛЬ ЗА РІВНЕМ ЕНЕРГОСПОЖИВАННЯ НА ОСНОВІ МОДИФІКОВАНОГО АЛГОРИТМУ K-MEANS

У цій роботі досліджується кластеризація будівель із застосуванням модифікованого алгоритму K-means, який передбачає розпаралелення обчислень і вибір початкових центроїдів на основі середніх значень. Запропонований підхід дозволяє класифікувати будівлі за рівнем енергоспоживання та забезпечує точнішу сегментацію споживачів електроенергії. Проведено експериментальне дослідження ефективності алгоритму при різних кількості процесів. Результати демонструють, що використання чотирьох процесорів забезпечує оптимальне пришвидшення виконання, що особливо актуально для систем моніторингу концепції розумного міста. Розпаралелення алгоритму K-means дозволяє значно зменшити час обробки великих обсягів даних, що робить його перспективним для прогнозування споживання електроенергії, оптимізації ресурсів та розробки програм енергозбереження.

Ключові слова: кластеризація, паралельні обчислення, енергоспоживання, аналіз даних, оптимізація, прогнозування, центроїди.

HENTOSH LESIA,**LEVKOYCH ROMAN**

Lviv Polytechnic National University

A PARALLEL APPROACH TO BUILDING CLUSTERING BASED ON ENERGY CONSUMPTION USING A MODIFIED K-MEANS ALGORITHM

This paper explores the clustering of buildings using a modified K-means clustering algorithm that incorporates parallelized computations and selects initial centroids based on mean values. The proposed approach enables the classification of buildings according to their energy consumption levels, ensuring more accurate segmentation of electricity consumers. By leveraging parallel processing, the algorithm significantly reduces computational complexity, making it more efficient for large-scale datasets. An experimental study was conducted to evaluate the algorithm's performance with different numbers of processes. The results indicate that the highest speedup was achieved using four processors, which corresponds to the number of available processing units in the system. This speedup reached up to 3.8–3.85 times compared to the sequential implementation, demonstrating the effectiveness of parallelization in handling large volumes of energy consumption data. The application of the parallelized K-means clustering algorithm in the context of energy consumption analysis offers multiple advantages. It facilitates the identification of distinct energy consumption patterns, which can be useful for developing energy-saving programs and optimizing power distribution. Furthermore, the ability to process energy data more frequently – potentially on a daily basis rather than weekly – enables more precise monitoring and forecasting of electricity usage. The findings suggest that parallel implementations of clustering algorithms are a promising direction for Smart City technologies. They can enhance the efficiency of urban energy management systems, contribute to the reduction of energy consumption, and improve the overall sustainability of modern cities. By supporting real-time analysis and decision-making, such algorithms pave the way for more adaptive and intelligent energy infrastructure.

Keywords: clustering, parallel computing, energy consumption, data analysis, optimization, prediction, centroids.

Постановка задачі

Зростаючий обсяг даних про енергоспоживання потребує ефективних методів їхнього аналізу для покращення управління енергоресурсами та підвищення енергоефективності будівель [1]. Класичний алгоритм K-means має обмеження, пов'язані з вибором початкових центроїдів та обчислювальною складністю при роботі з великими наборами даних. Запропонована модифікація з використанням середнього значення для ініціалізації центроїдів та розпаралелення процесу кластеризації дозволяє підвищити швидкість і точність алгоритму. Це сприятиме більш ефективному аналізу споживання електроенергії, що може бути корисним для оптимізації енергетичних ресурсів, розробки стратегій економії та планування майбутніх енергетичних політик.

Метою цього дослідження є кластеризація даних про споживання електроенергії будинками шляхом розробки та застосування модифікованого алгоритму кластиризації K-means із розпаралеленням процесів та використанням середнього значення певного проміжку даних для ініціалізації центроїдів. Очікуваним результатом є ефективне розбиття об'єктів (споживання електроенергії будинками в певні дні) на кластери з подібними характеристиками, що дозволить виявити закономірності у споживанні енергії.

Далі сформулюємо постановку розглядуваної задачі. Дано множину даних про споживання електроенергії будинками: $M = \{m_1, m_2, \dots, m_n\}$, де кожна точка m_i – представлена тривимірним вектором координат (x, y, z) , де x – ідентифікатор будинку, y – день року, z – кількість спожитої

електроенергії у цей день.

Необхідно здійснити кластеризацію множини M на k кластерів за допомогою модифікованого алгоритму K -means. При цьому слід врахувати наступні обмеження та умови. Центроїди кластерів ініціалізуються не випадковим чином, а на основі середнього значення певного проміжку даних, що дозволяє зменшити вплив початкового вибору на результати кластеризації. Алгоритм має бути розпаралелений таким чином, щоб процес призначення точок даних до найближчих кластерів виконувався паралельно, що дозволить підвищити швидкість обчислень. Необхідно оцінити швидкодію алгоритму, зокрема: час виконання при різній кількості процесів ітерацій, коефіцієнт прискорення (очікується, що він буде більшим за 2.0), ефективність паралельних обчислень.

При цьому очікується, що запропонований підхід забезпечить більш стабільну та швидку кластеризацію порівняно з класичним K -means за рахунок оптимізації вибору початкових центроїдів та використання паралельних обчислень. Завдяки цьому можливо зменшити кількість ітерацій, необхідних для збіжності алгоритму, а також підвищити точність отриманих кластерів шляхом мінімізації впливу випадкового вибору початкових центроїдів. Додатково, розпаралелення дозволить скоротити час виконання кластеризації на великих наборах даних, що робить підхід перспективним для застосування в задачах аналізу енергоспоживання в реальних умовах.

Наукова новизна цього дослідження полягає у розробці модифікованого алгоритму K -means, який поєднує оптимізований вибір початкових центроїдів на основі середніх значень певного проміжку даних із розпаралеленням процесу призначення об'єктів до кластерів. Такий підхід зменшує чутливість алгоритму до початкових умов, скорочує кількість ітерацій для збіжності та підвищує стабільність кластеризації. Використання паралельних обчислень значно пришвидшує процес аналізу великих обсягів даних, що робить запропонований метод ефективним для дослідження споживання електроенергії. Дослідження також включає оцінку обчислювальної складності, прискорення та ефективності запропонованого підходу, що дозволяє визначити його практичну доцільність у сфері енергетичного моніторингу.

Аналіз останніх джерел

Розвиток інформаційних технологій є ключовим фактором у формуванні концепції розумного міста, зокрема завдяки використанню великих обсягів даних і розподілених обчислень для оптимізації міської інфраструктури [2]. Паралельні алгоритми відіграють важливу роль у прискоренні обробки даних, що стає критично важливим у зв'язку зі зростанням їх обсягу [3]. Одним із перспективних напрямів є кластеризація будівель за щоденним енергоспоживанням, що сприяє виявленню закономірностей та підвищенню ефективності використання електроенергії.

У [4] автори проаналізували ефективність використання паралельних алгоритмів у сфері аналізу даних у розумних містах. Зокрема, вони розглянули розподілені бази даних, паралельні алгоритми машинного навчання та розподілені алгоритми кластеризації. Одним із висновків дослідження є ефективність випадкового вибору центроїдів, проте в межах розробленого підходу доцільніше використовувати центроїди, визначені як середнє значення координат точок. Це пояснюється тим, що при паралельній обробці кожен процес працює з окремим піднабором даних, і використання середніх значень сприяє більш коректному узагальненню розподілу даних у кожному процесі.

У роботі [5] досліджено різні аспекти застосування паралельних і розподілених алгоритмів у контексті розумних міст. Автор розглядає техніки розподілених обчислень, зокрема, системи розподілених баз даних, паралельні алгоритми та технології розподіленого машинного навчання. Зокрема, алгоритми кластеризації вивчалися переважно у їхніх стандартних реалізаціях, де підвищення ефективності досягалося завдяки попередньому розподілу даних у розподілених системах. У даній роботі ефективність кластеризації підвищується безпосередньо за рахунок розпаралелення самого алгоритму, що дозволяє теоретично досягти прискорення у L разів (де L – кількість потоків). Водночас, на практиці ефективність буде нижчою через накладні витрати, пов'язані з ініціалізацією та завершенням обчислювальних процесів.

У [6] автори дослідили різні методи кластеризації у розумних містах з використанням штучного інтелекту (AI). Вони розглянули ефективність різних підходів, орієнтуючись на якість та продуктивність алгоритмів. Проте використання AI-методів може бути обмежене якістю навчальних даних: у разі недостатньої або нерепрезентативної вибірки модель може демонструвати недостатню узагальнювальну здатність або, навпаки, перенавчання. У межах цієї роботи розглядається алгоритмічний підхід, що не передбачає використання AI-моделей, що дозволяє забезпечити більш надійний аналіз споживання електроенергії незалежно від специфіки навчальних даних.

Одним з важливих завдань машинного навчання в розумному місті є кластеризація, що дозволяє розподіляти об'єкти на групи за певними критеріями. Одним із найпоширеніших методів є алгоритм K -means, який широко використовується для аналізу споживання електроенергії будівлями з метою підвищення ефективності її використання. Зокрема, у статті [7] зазначено, що кластерний аналіз сприяє ефективному управлінню енергоспоживанням у містах, дозволяючи виявити ключові характеристики кожного класу споживачів. Використання такого підходу сприяє зниженню витрат на енергію та мінімізації негативного впливу на довкілля. Однак у цьому дослідженні не розглядається можливість розпаралелення алгоритмів для обробки великих потоків даних.

У [8] розглядається використання кластерного аналізу для класифікації споживачів електроенергії за тижневими витратами. Хоча цей підхід є ефективним, точність аналізу можна покращити, використовуючи дані про щоденні витрати, що значно збільшить обсяг даних. Використання розпаралеленої версії алгоритму K-means на багатоядерних обчислювальних системах дозволяє зменшити час обробки таких великих обсягів інформації.

Дослідження [9] пропонує покращення алгоритму K-means шляхом збереження проміжних даних між ітераціями. Однак цей підхід є менш ефективним для паралельних обчислень, оскільки потребує синхронізації потоків, що може знизити продуктивність через необхідність блокування доступу до спільних ресурсів.

У [10] досліджено ефективність паралельного виконання K-means на графічних процесорах. Виявлено, що використання графічних процесорів може скоротити час виконання алгоритму в 25 разів порівняно з традиційними методами на центральних процесорах. Проте існують певні обмеження, зокрема обсяг пам'яті, обробка розріджених даних та масштабованість ітеративного процесу. Використання паралельного виконання алгоритму дозволяє подолати частину цих проблем.

Таким чином, алгоритм K-means є ефективним інструментом кластеризації великих обсягів даних у розумному місті, особливо для аналізу енергоспоживання. Його ефективність можна підвищити шляхом розпаралелення обчислень, що дозволяє значно зменшити час виконання та забезпечити масштабованість. Водночас слід враховувати певні обмеження алгоритму, зокрема його чутливість до початкового вибору центроїдів та лінійне зростання складності обчислень із розміром даних. Запропоновані підходи до розпаралелення сприяють підвищенню продуктивності та надійності аналізу великих обсягів інформації у сфері розумного міста.

Виклад основного матеріалу

Кластеризація є важливим інструментом аналізу даних, який дозволяє групувати схожі об'єкти в окремі класи. Одним із поширених алгоритмів кластеризації є K-means [11], що використовується для розподілу даних на k кластерів шляхом ітеративного коригування центроїдів. У даній роботі розглянуто паралельний підхід до кластеризації даних споживання електроенергії будинками, що дозволяє зменшити обчислювальну складність за рахунок багатопроцесорної обробки.

Запропонований підхід передбачає наступні етапи:

1. Визначення кількості кластерів k .
2. Розподіл множини точок m на k рівних підмножин $\{f_1, f_2, \dots, f_k\}$ з урахуванням вирівнювання залишкових елементів
3. Обчислення початкових координат центроїдів u_j (x', y', z') за формулами:

$$x' = \frac{1}{n} \sum_{i=1}^n x_i, \quad y' = \frac{1}{n} \sum_{i=1}^n y_i, \quad z' = \frac{1}{n} \sum_{i=1}^n z_i,$$
де x_i, y_i, z_i – це координати кожної точки в цій підмножині. Середні координати x', y', z' обчислюються як середнє арифметичне всіх точок у відповідній підмножині.
4. Розподіл точок m на l частин і створення l паралельних процесів, які для кожної точки обчислюють відстань D до всіх центроїдів та призначають точку найближчому кластеру.
5. Оновлення координат центроїдів за середнім значенням координат точок, що належать до кластеру.
6. Повторення кроків 4-5 до збіжності центроїдів (критерій – малий приріст координат між ітераціями) або досягнення максимальної кількості ітерацій t .

Для реалізації паралельного підходу використовується бібліотека multiprocessing у Python. Вона забезпечує розподіл завдань між незалежними процесами, що працюють на окремих ядрах процесора. Основні переваги multiprocessing:

- Зменшення часу виконання за рахунок одночасної обробки різних частин набору даних.
- Висока продуктивність у порівнянні з однопотокowymi підходами.
- Кросплатформна підтримка.

Розрахункова складність алгоритму становить $O(nkt)$, де n – кількість точок даних, k – кількість кластерів, t – кількість ітерацій. За рахунок розпаралелення на l процесів досягається зменшення складності до $O(nkt/l)$.

Для реалізації алгоритму використано:

- *numpy* для обчислень над масивами даних,
- *pandas* для зручної роботи з табличними даними,
- *multiprocessing* для розпаралелення обчислень.

Функції *numpy*, такі як *np.zeros* (ініціалізація масивів), *np.inf* (початкове значення мінімальної відстані) та *np.sqrt* (обчислення евклідової відстані), забезпечують ефективність обчислень. *pandas* використовується для читування та обробки вхідних даних.

У ході експериментів було використано датасет [12] та проведено кластеризацію будівель за їхнім щоденним енергоспоживанням за допомогою модифікованого алгоритму K-means із розпаралеленням обчислень та без нього. Аналіз отриманих кластерів показав рівномірний розподіл об'єктів між групами, що свідчить про стабільність і точність роботи алгоритму незалежно від обраного підходу. Результати кластеризації для п'яти кластерів наведено в Таблиці 1.

Таблиця 1

Розподіл об'єктів по кластерах		
№ Кластеру	Кількість об'єктів (розпаралелений алгоритм)	Кількість об'єктів (звичайний алгоритм)
0	7346	7388
1	6924	6812
2	7681	7733
3	7463	7523
4	7086	7044

Обидва варіанти алгоритму, представлені у Таблиці 1 демонструють рівномірний розподіл об'єктів між кластерами, що свідчить про їхню ефективність. Окрім аналізу розподілу кластерів, було проведено оцінку швидкодії алгоритму залежно від кількості процесів [13]. Час виконання алгоритму оцінювався при різній кількості процесів та ітерацій. Результати наведено у Таблицях 2–4.

Таблиця 2

Час виконання, пришвидшення та ефективність при кількості ітерацій 1

Кількість процесів	Час виконання (с)	Пришвидшення	Ефективність
1	24.743	1.000	1.000
2	19.841	1.247	0.311
3	15.245	1.623	0.405
4	13.048	1.896	0.474
8	12.657	1.954	0.488
16	12.172	2.032	0.508

Таблиця 3

Час виконання, пришвидшення та ефективність при кількості ітерацій 2

Кількість процесів	Час виконання (с)	Пришвидшення	Ефективність
1	39.819	1.000	1.000
2	28.982	1.373	0.343
3	19.342	2.058	0.514
4	12.137	3.280	0.820
8	11.856	3.358	0.839
16	12.047	3.305	0.826

Таблиця 4

Час виконання, пришвидшення та ефективність при кількості ітерацій 10

Кількість процесів	Час виконання (с)	Пришвидшення	Ефективність
1	211.563	1.000	1.000
2	146.773	1.441	0.360
3	82.521	2.564	0.641
4	55.339	3.823	0.956
8	54.937	3.851	0.962
16	55.417	3.817	0.954

Як видно з результатів, пришвидшення алгоритму відбувається пропорційно збільшенню кількості процесів, але після певного рівня (4 процеси) спостерігається спад ефективності. Це пов'язано з обмеженням апаратних ресурсів і накладними витратами на управління потоками. Найкращі результати досягнуто при використанні 4 процесів, що дозволило пришвидшити обчислення у 3.823 рази при 10 ітераціях.

Зі збільшенням кількості ітерацій пришвидшення алгоритму зростає. При 10 ітераціях максимальне пришвидшення сягнуло 3.85 разів у порівнянні з виконанням у один потік, що свідчить про ефективність запропонованої модифікації алгоритму.

Таким чином, модифікований K-means із розпаралеленням демонструє хорошу ефективність при великій кількості ітерацій. Виявлено, що найбільший приріст швидкості спостерігається при переході від 1 до 4 процесів, після чого ефективність стабілізується. Це означає, що запропонований підхід є особливо корисним для обробки великих обсягів даних із багаторазовими ітераціями алгоритму кластеризації.

Висновки

У цій роботі було розглянуто проблему кластеризації будинків за даними їхнього енергоспоживання та її вирішення за допомогою розпаралеленої версії алгоритму K-means із використанням центроїдів, базуючись на середньому значенні даних. Запропонований підхід дозволив ефективно розподілити будинки на кластери за схожістю енергоспоживання, що може сприяти оптимізації розподілу енергоресурсів у межах концепції розумного міста.

Проведений аналіз ефективності алгоритму при різній кількості процесів показав, що найкращі результати за швидкістю та пришвидшенням виконання досягнуто при використанні чотирьох процесорів. Це пояснюється балансом між розподілом обчислювального навантаження та комунікаційними витратами в паралельній обробці. Досягнуте пришвидшення в 3.8-3.85 разів свідчить про перспективність використання таких підходів для обробки великих обсягів даних у системах моніторингу енергоспоживання.

Застосування розпаралеленого алгоритму K-means у сфері енергоменеджменту дозволяє визначати основні патерни споживання електроенергії, що може бути корисним для розробки стратегій енергозбереження, прогнозування навантаження на енергосистему та оптимізації тарифікації. Крім того, завдяки пришвидшенню обчислень, можливим є перехід від щотижневого до щоденного аналізу даних, що підвищує оперативність прийняття управлінських рішень.

Подальші дослідження можуть бути спрямовані на вдосконалення алгоритму шляхом адаптивного вибору початкових центрів кластерів для підвищення точності кластеризації та зменшення кількості ітерацій. Перспективним напрямом є дослідження ефективності використання інших методів кластеризації, зокрема ієрархічних або алгоритмів, заснованих на щільності [14], у контексті аналізу енергоспоживання. Іншою важливою перспективою є інтеграція алгоритмів кластеризації з методами глибокого навчання для більш точної та автоматизованої сегментації даних. Додатково, дослідження може бути розширене на інші аспекти розумного міста, зокрема оптимізацію розподілу навантажень у міських енергомережах та аналіз ефективності відновлюваних джерел енергії в різних категоріях будівель [15].

Таким чином, результати роботи підтверджують ефективність використання розпаралелених алгоритмів кластеризації для аналізу енергоспоживання в рамках концепції розумного міста. Подальший розвиток цих методів може сприяти підвищенню ефективності управління міськими ресурсами, зниженню витрат енергії та покращенню якості життя мешканців урбанізованих територій.

Література

1. Shakhovska N. Innovative machine learning approaches for indoor air temperature forecasting in smart infrastructure. / N. Shakhovska, L. Mochurad, R. Caro, and S. Argyroudis. // *Sci. Rep.* – 2025. – vol. 15, no. 1. – P. 1-15.
2. Mochurad L. A fast parallelized DBSCAN algorithm based on OpenMp for detection of criminals on streaming services. / L. Mochurad, A. Sydor, and O. Ratinskiy. // *Front. Big Data.* – 2023. – vol. 6. – P. 1-12.
3. [Popova S. Application of the Smart House System for Reconstruction of Residential Buildings from an Obsolete Housing Stock. / S. Popova and I. Izonin. // *Smart Cities.* – 2022. – vol. 6, no. 1. – P. 57–71.
4. Xiao W. A Survey of Parallel Clustering Algorithms Based on Spark. / W. Xiao and J. Hu. // *Sci. Program.* – 2020. – vol. 2020. – P. 1–12.
5. Nasiri H. A Survey of Distributed Stream Processing Systems for Smart City Data Analytics. / H. Nasiri, S. Nasehi, and M. Goudarzi. // *Proceedings of the international conference on smart cities and internet of things, Mashhad Iran: ACM.* – 2018. – P. 1–7.
6. Tharwat M. Clustering Techniques for Smart Cities: An Artificial Intelligence Perspective. / M. Tharwat and A. Khattab. // *Smart Cities: A Data Analytics Perspective*, M. A. Khan, F. Algarni, and M. T. Quasim, Eds., in *Lecture Notes in Intelligent Transportation and Infrastructure*, Cham: Springer International Publishing. – 2021. – P. 113–134.
7. Sayah Z. An intelligent system for energy management in smart cities based on big data and ontology. / Z. Sayah, O. Kazar, B. Lejdel, A. Laouid, and A. Ghenabzia. // *Smart Sustain. Built Environ.* – 2021. – vol. 10, no. 2. P. 169–192.
8. Si C. Electric Load Clustering in Smart Grid: Methodologies, Applications, and Future Trends. / C. Si, S. Xu, C. Wan, D. Chen, W. Cui, and J. Zhao. // *J. Mod. Power Syst. Clean Energy.* – 2021. – vol. 9, no. 2. – P. 237–252.
9. Na S. Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. / S. Na, L. Xumin, and G. Yong. // *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*, Jian, China: IEEE. – 2010. – P. 63–67.
10. Li M. Large scale K-means clustering using GPUs. / M. Li, E. Frank, and B. Pfahringer. // *Data Min. Knowl. Discov.* – 2023. – vol. 37, no. 1. – P. 67–109.
11. Oti E. U. Comprehensive Review of K-Means Clustering Algorithms. / E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe // *Int. J. Adv. Sci. Res. Eng.* – 2021. – vol. 07, no. 08. – P. 64–69.

12. "Energy_data." Feb. 11, 2025. [Online]. Available: <https://drive.google.com/file/d/1pAAshEbaDNWVQotXZaYrdzoKqY3yJ6Sr/view>
13. Mochurad L. Implementation and analysis of a parallel kalman filter algorithm for lidar localization based on CUDA technology. / L. Mochurad // *Front. Robot. AI*. – 2024. – vol. 11. – P. 1-11.
14. Bhuyan R. A Survey of Some Density Based Clustering Techniques. / Rupanka Bhuyan and Samarjeet Borah. // *arXiv preprint arXiv:2306.09256*. – 2013. – P. 1-4.
15. Alsayaydeh J. A. J. Development of Vehicle Ignition Using Fingerprint / J. A. J. Alsayaydeh, W. A. Indra, A. W. Y. Khang, V. Shkaruplyo, and D. A. P. P. Jkatisan // *ARNP Journal of Engineering and Applied Sciences*. – 2019. – vol. 14, no. 23. – P. 4045-4053.

References

1. Shakhovska N. Innovative machine learning approaches for indoor air temperature forecasting in smart infrastructure. / N. Shakhovska, L. Mochurad, R. Caro, and S. Argyroudis. // *Sci. Rep.* – 2025. – vol. 15, no. 1. – P. 1-15.
2. Mochurad L. A fast parallelized DBSCAN algorithm based on OpenMp for detection of criminals on streaming services. / L. Mochurad, A. Sydor, and O. Ratinskiy. // *Front. Big Data*. – 2023. – vol. 6. – P. 1-12.
3. [Popova S. Application of the Smart House System for Reconstruction of Residential Buildings from an Obsolete Housing Stock. / S. Popova and I. Izonin. // *Smart Cities*. – 2022. – vol. 6, no. 1. – P. 57-71.
4. Xiao W. A Survey of Parallel Clustering Algorithms Based on Spark. / W. Xiao and J. Hu. // *Sci. Program*. – 2020. – vol. 2020. – P. 1-12.
5. Nasiri H. A Survey of Distributed Stream Processing Systems for Smart City Data Analytics. / H. Nasiri, S. Nasehi, and M. Goudarzi. // *Proceedings of the international conference on smart cities and internet of things, Mashhad Iran: ACM*. – 2018. – P. 1-7.
6. Tharwat M. Clustering Techniques for Smart Cities: An Artificial Intelligence Perspective. / M. Tharwat and A. Khattab. // *Smart Cities: A Data Analytics Perspective*, M. A. Khan, F. Algarni, and M. T. Quasim, Eds., in *Lecture Notes in Intelligent Transportation and Infrastructure*, Cham: Springer International Publishing. – 2021. – P. 113-134.
7. Sayah Z. An intelligent system for energy management in smart cities based on big data and ontology. / Z. Sayah, O. Kazar, B. Lejdel, A. Laouid, and A. Ghenabzia. // *Smart Sustain. Built Environ*. – 2021. – vol. 10, no. 2. P. 169-192.
8. Si C. Electric Load Clustering in Smart Grid: Methodologies, Applications, and Future Trends. / C. Si, S. Xu, C. Wan, D. Chen, W. Cui, and J. Zhao. // *J. Mod. Power Syst. Clean Energy*. – 2021. – vol. 9, no. 2. – P. 237-252.
9. Na S. Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. / S. Na, L. Xumin, and G. Yong. // *2010 Third International Symposium on Intelligent Information Technology and Security Informatics, Jian, China: IEEE*. – 2010. – P. 63-67.
10. Li M. Large scale K-means clustering using GPUs. / M. Li, E. Frank, and B. Pfahringer. // *Data Min. Knowl. Discov*. – 2023. – vol. 37, no. 1. – P. 67-109.
11. Oti E. U. Comprehensive Review of K-Means Clustering Algorithms. / E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe // *Int. J. Adv. Sci. Res. Eng*. – 2021. – vol. 07, no. 08. – P. 64-69.
12. "Energy_data." Feb. 11, 2025. [Online]. Available: <https://drive.google.com/file/d/1pAAshEbaDNWVQotXZaYrdzoKqY3yJ6Sr/view>
13. Mochurad L. Implementation and analysis of a parallel kalman filter algorithm for lidar localization based on CUDA technology. / L. Mochurad // *Front. Robot. AI*. – 2024. – vol. 11. – P. 1-11.
14. Bhuyan R. A Survey of Some Density Based Clustering Techniques. / Rupanka Bhuyan and Samarjeet Borah. // *arXiv preprint arXiv:2306.09256*. – 2013. – P. 1-4.
15. Alsayaydeh J. A. J. Development of Vehicle Ignition Using Fingerprint / J. A. J. Alsayaydeh, W. A. Indra, A. W. Y. Khang, V. Shkaruplyo, and D. A. P. P. Jkatisan // *ARNP Journal of Engineering and Applied Sciences*. – 2019. – vol. 14, no. 23. – P. 4045-4053.