

DOI: <https://doi.org/10.36910/6775-2524-0560-2025-60-22>

УДК 004:85

Лось Владислав Юрійович, аспірант

<https://orcid.org/0009-0009-3948-4392>

Черкаський національний університет імені Богдана Хмельницького, м. Черкаси, Україна

ІДЕНТИФІКАЦІЯ РЕПРЕЗЕНТАТИВНИХ ПАТЕРНІВ СПОЖИВАННЯ ЕНЕРГІЇ: РОЗРОБКА ТА ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДОЛОГІЙ

Лось В. Ю. Ідентифікація репрезентативних патернів споживання енергії: розробка та порівняльний аналіз методологій. У статті представлено розробку та емпіричну оцінку методології ідентифікації найбільш репрезентативного добового патерну споживання електроенергії, що є ключовим завданням для оптимізації інтелектуальних енергетичних систем (Smart Grid), прогнозування навантаження та управління попитом. Запропонований підхід фокусується на оптимізації попередньої обробки даних та використанні багатofакторної метрики подібності. Він включає двоетапну адаптивну фільтрацію добових профілів на основі пікової та середньої потужності, що дозволяє ефективно відсіювати аномальні та неактивні дні завдяки динамічним порогам, які розраховуються на основі статистичних індикаторів (медіана, перцентилі), адаптуючись до індивідуальних особливостей споживання. Для вибору репрезентативного патерну розроблено композитну метрику, яка комплексно інтегрує схожість за формою, рівнем потужності та часовими характеристиками активних періодів. Наукова новизна полягає у гібридній фільтрації та прямому виборі медоїда без повноцінної кластеризації, що спрощує обчислення та усуває необхідність попереднього визначення кількості кластерів. Методологія забезпечує швидке та обґрунтоване визначення найбільш характерного профілю споживання, який може слугувати надійною базовою моделлю (baseline) для виявлення аномалій, сегментації споживачів або генерації синтетичних даних навантаження. Результати тестування на реальних даних з датасету "ECO data set" підтверджують ефективність підходу для різних електроприладів. Дане дослідження робить значний внесок у розвиток методів управління енергоспоживанням. Перспективи включають інтеграцію "еластичних" метрик, формалізовану кластеризацію та розробку складних систем виявлення аномалій.

Ключові слова: патерни споживання енергії, Smart Grid, кластеризація, виявлення аномалій, композитна метрика подібності, медоїд.

Los V. Identification of representative energy consumption patterns: development and comparative analysis of methodologies. This paper presents the development and empirical evaluation of a methodology for identifying the most representative daily energy consumption pattern, a key task for optimizing Smart Grid systems, load forecasting, and demand management. The proposed approach focuses on optimizing data preprocessing and utilizing a multi-factor similarity metric. It includes a two-stage adaptive filtering of daily profiles based on peak and average power, which allows for the effective exclusion of anomalous and inactive days through dynamic thresholds calculated based on statistical indicators (median, percentiles), adapting to individual consumption characteristics. To select the representative pattern, a composite metric was developed that comprehensively integrates similarities in shape, power level, and the temporal characteristics of active periods. The scientific novelty lies in the hybrid filtering and direct medoid selection without full-fledged clustering, which simplifies computation and eliminates the need to predefine the number of clusters. The methodology provides a fast and justified determination of the most typical consumption profile, which can serve as a reliable baseline model for anomaly detection, consumer segmentation, or synthetic load data generation. Testing results on real data from the "ECO data set" confirm the approach's effectiveness for various electrical appliances. This research makes a significant contribution to the advancement of energy management methods. Prospects include the integration of "elastic" metrics, formalized clustering, and the development of sophisticated anomaly detection systems.

Keywords: energy consumption patterns, Smart Grid, clustering, anomaly detection, composite similarity metric, medoid.

Постановка проблеми. В умовах глобальної цифровізації та зростаючого впровадження інтелектуальних енергетичних систем (Smart Grids), аналіз даних про споживання енергії набуває критичного значення. Ефективне управління попитом, оптимізація генерації, впровадження енергоефективних технологій та прогнозування навантаження неможливі без глибокого розуміння типових поведінкових патернів споживачів. Зокрема, ідентифікація репрезентативних патернів добового споживання дозволяє створювати базові моделі для прогнозування, виявляти аномалії, проводити сегментацію споживачів та оцінювати ефективність впровадження нових технологій. Ця проблема має прямий зв'язок із важливими науковими завданнями у сфері Data Mining, Machine Learning та Smart Grid Optimization, а також з практичними завданнями для енергетичних компаній, постачальників послуг та кінцевих споживачів. Точне виявлення таких патернів є фундаментом для прийняття обґрунтованих рішень щодо інфраструктури, тарифної політики та стратегій зниження енергоспоживання.

Аналіз останніх досліджень та публікацій. Проблема ідентифікації патернів споживання енергії активно досліджується протягом останніх десятиліть. Значний обсяг робіт зосереджений на застосуванні алгоритмів кластеризації для групування добових профілів навантаження. Класичні методи, такі як K-Means, широко використовуються завдяки їхній простоті та обчислювальній

ефективності, проте вони чутливі до викидів і не завжди ефективні для виявлення складної структури даних. Для подолання цих обмежень були запропоновані модифікації, наприклад, K-Medoids, що обирає центроїдами реальні точки даних, роблячи його більш стійким до аномалій [1].

М'які методи кластеризації, такі як Fuzzy C-Means та Gaussian Mixture Models (GMM), дозволяють профілям належати до кількох кластерів з різною мірою ймовірності, що краще відображає перехідні або змішані патерни поведінки споживачів [2, 3].

Окрім кластеризації, значну увагу приділяють вилученню ознак (feature engineering) з часових рядів. Це можуть бути статистичні показники (середнє, пік, дисперсія), сезонні агрегації, або більш складні характеристики, отримані за допомогою перетворення вейвлетів (Discrete Wavelet Transform - DWT) [4] або емпіричної модальної декомпозиції (Empirical Mode Decomposition - EMD) [5]. Ці методи дозволяють розкласти сигнал на компоненти, виділяючи як базове навантаження, так і динамічні коливання (мінливість).

Вибір метрик подібності також є ключовим. Хоча Евклідова відстань є найпростішою, вона не справляється з часовими зсувами або деформаціями. Більш надійними є "еластичні" метрики, такі як Dynamic Time Warping (DTW) [6] та Move Split Merge (MSM) [7], які дозволяють деформувати часову вісь для пошуку оптимального зіставлення патернів. Деякі дослідження також використовують кореляційні метрики (Pearson correlation) або косинусну подібність для оцінки форми патернів незалежно від їхньої амплітуди.

Окремий напрямок – це виявлення аномалій (anomaly detection), що використовується для попередньої фільтрації даних, вилучення шумів та нетипових днів, які можуть спотворити виявлення репрезентативних патернів [8].

Виділення невирішених раніше частин проблеми. Існуючі дослідження пропонують потужні інструменти, проте є кілька аспектів, які вимагають подальшого вдосконалення, зокрема у контексті автоматизованого та адаптивного виявлення патернів для конкретних пристроїв або домогосподарств без значного ручного втручання:

- Адаптивність фільтрації: Більшість методів фільтрації активних днів або виявлення аномалій вимагають попереднього налаштування порогових значень або базуються на глобальних статистичних показниках, які не завжди враховують індивідуальні особливості пристроїв чи змінну природу споживання. Потреба в більш адаптивних, можливо, самонавчальних механізмах фільтрації є очевидною.

- Гібридні метрики подібності: Об'єднання різних аспектів подібності (форми, амплітуди, часових характеристик) в єдину метрику є перспективним, але вимагає обґрунтованого підбору ваг та вибору компонентів, що враховують як глобальну схожість, так і локальні особливості.

- Інтерпретованість "репрезентативного" патерну: Хоча K-Medoids обирає "реальний" профіль як репрезентативний, для багатьох застосувань важливо не просто мати один типовий профіль, а й розуміти, які саме характеристики роблять його таким, та як він співвідноситься з іншими варіаціями.

- Обробка "відносно неактивних" днів: Визначення "неактивного" дня може бути складним. Пристрій може мати пік активності, але низьке середнє споживання (наприклад, короткочасне потужне включення). Поточні методи часто жорстко фільтрують такі дні, що може призвести до втрати цінної інформації про "типові" короткочасні режими роботи.

Означена стаття присвячена розробці та детальному аналізу прагматичного підходу до ідентифікації репрезентативного патерну, який враховує як пікову, так і середню потужність для попередньої фільтрації, а також використовує гібридну метрику подібності для вибору єдиного найбільш типового добового профілю.

Формулювання мети дослідження. Метою даного дослідження є розробка та емпірична оцінка методології виявлення найбільш репрезентативного добового патерну споживання енергії на основі аналізу історичних даних, з акцентом на ефективній попередній фільтрації нерелевантних днів та використанні багатofакторної метрики подібності. Завдання включає:

- Створення адаптивної системи фільтрації добових профілів за критеріями пікової та середньої потужності.

- Розробку композитної метрики подібності, що інтегрує схожість за формою, рівнем потужності та часом активних періодів.

- Програмну реалізацію та апробацію запропонованого підходу на реальних даних споживання.
- Порівняльний аналіз отриманих результатів та архітектури методології з існуючими академічними підходами.

Виклад основного матеріалу дослідження. Запропонована методологія призначена для ідентифікації єдиного "найбільш репрезентативного" добового профілю з часового ряду вимірювань потужності $P = p_{t=1}^N$, де N загальна кількість вимірювань. Вхідний часовий ряд даних про споживання енергії розбивається на окремі добові патерни (1):

$$D_i = p_{i,j=1}^S \quad (1)$$

Кожен такий патерн складається з фіксованої кількості вимірювань S (наприклад, 96 слотів для 15-хвилинних інтервалів). На етапі попередньої обробки неповні добові патерни ігноруються, що забезпечує якість подальшого аналізу. Загальна блок-схема алгоритму зображена на (Рис.1).

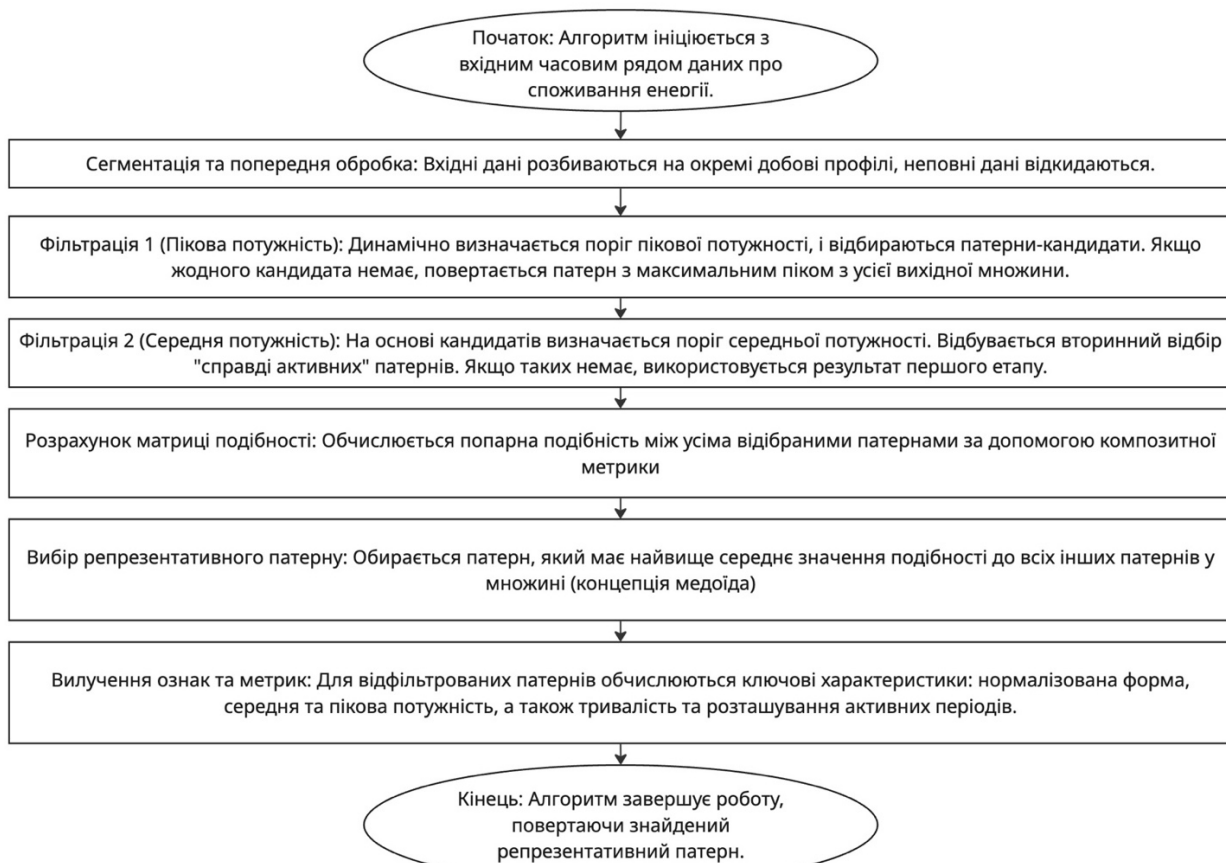


Рис.1. Блок-схема алгоритму

Для відбору днів, що справді відображають "використання пристрою", застосовується двоетапна адаптивна фільтрація:

- Етап 1: Фільтрація за піковою потужністю. Спочатку відбираються "кандидати на активні патерни" на основі їхнього пікового споживання. Порогове значення для пікової потужності (2) T_{peak} визначається динамічно, враховуючи медіану (P_{peak}^{med}) та 95-й перцентиль (P_{peak}^{95}) пікових потужностей з усього набору даних.

$$T_{peak} = \max(10.0, 0.5 \cdot P_{peak}^{med}, 0.1 \cdot P_{peak}^{95}) \quad (2)$$

Це дозволяє фільтру адаптуватися до загального рівня потужності у даних, уникаючи жорстко фіксованих порогів.

- **Етап 2: Фільтрація за середньою потужністю ("істотне використання").** Другий етап посилює фільтрацію, вимагаючи не лише піка, а й стійкого використання. Для цього розраховується порогове значення T_{mean} на основі 25-го перцентиля середніх значень потужності для патернів, які пройшли перший етап. Це допомагає відрізнити короточасні викиди від днів зі справжнім, тривалим споживанням. Якщо після цієї фільтрації не залишається "справді активних" патернів, для подальшого аналізу використовується результат першого етапу.

Для кожного відфільтрованого патерну обчислюються ключові ознаки, що використовуються для порівняння:

- **Нормалізований патерн:** кожне значення патерну масштабується до діапазону $[0,1]$. Це дозволяє порівнювати форму патернів, ігноруючи їхню абсолютну амплітуду.
- **Середня та пікова потужність:** абсолютні значення споживання, що характеризують масштаб використання.
- **Активні періоди:** визначаються як часові слоти, де потужність перевищує певний динамічний поріг. Зберігаються їхня тривалість та початкові моменти, що дає уявлення про часову синхронізацію використання.

Вибір найбільш репрезентативного патерну ґрунтується на концепції медоїда – патерну, який є найближчим до всіх інших патернів у відібраній групі. Застосовується багатовимірний композитний метричний подібності (3) $Sim(D_a, D_b)$, що складається з трьох компонентів з фіксованими вагами ($w_1 = 0.4$ для форми, $w_2 = 0.2$ для рівня потужності, $w_3 = 0.4$ для активних періодів):

$$Sim(D_a, D_b) = w_1 \cdot Sim_{shape} + w_2 \cdot Sim_{level} + w_3 \cdot Sim_{active} \quad (3)$$

Компоненти подібності:

- **Подібність форми (Sim_{shape}):** Обчислюється на основі нормованої L1-відстані між нормалізованими патернами. Високе значення означає, що патерни мають схожий "силует".
- **Подібність за рівнем потужності (Sim_{level}):** Визначається як відношення мінімального середнього значення до максимального середнього значення потужності між двома патернами, показуючи наскільки близькі загальні рівні споживання.
- **Подібність активних періодів (Sim_{active}):** Комбінує індекс Жаккара для множин активних слотів та співвідношення загальних активних тривалостей. Це дає уявлення про схожість у часі та тривалості активного використання.

Для кожного патерну D_k у відфільтрованій множині обчислюється його середня подібність до всіх інших патернів. Репрезентативний патерн (4) D_{rep} вибирається як патерн з максимальною середньою подібністю:

$$D_{rep} = \arg \max_{D_k} \overline{Sim}(D_k) \quad (4)$$

Результати тестування. Для тестування запропонованої методології був використаний датасет "ECO data set (Electricity Consumption & Occupancy)" [9] з даними про споживання електроенергії домогосподарства #05. Цей датасет має в собі високочастотні данні спостережень споживання приладів за період з 27.06.12 по 31.01.13. Період аналізу охоплював 219 днів з інтервалом вимірювань агрегованих до 15 хвилинних інтервалів, що становить 96 слотів на добу. Аналіз проводиться для 7 різних електроприладів. Результати застосування алгоритму з проміжними значеннями які використовуються для пошуку серед історичних даних періоду який можна вважати репрезентативним зображено в Таблиці 1. Ця таблиця відображає основні характеристики споживання для кожного приладу, включаючи кількість днів з даними, кількість днів з фактичним використанням, відсоток використання, медіанні та 95-ті перцентилі пікових навантажень, максимальні піки та медіанні середньодобові значення. Це дозволяє оцінити особливості використання кожного приладу та його типові характеристики споживання, що є основою для застосування нашої адаптивної фільтрації.

Таблиця 1. Статистика використання приладів за тестовий період:

Appliance	Days with Data	Days with Usage	Usage %	Median Peak (W)	95th Percentile Peak (W)	Max Peak (W)	Median Daily Mean (W)	Days with Peak > 1W
Tablet	219	5	2.3%	8.6	10.7	15.0	6.7	218
Coffee machine	219	209	95.4%	1528.2	1564.5	1581.5	101.1	213
Fountain	219	70	32.0%	43.0	43.0	45.1	12.8	71
Microwave	219	62	28.3%	6.6	2648.7	2680.6	6.6	218
Fridge	219	218	99.5%	1301.1	1324.8	1358.4	251.3	218
Entertainment	219	189	86.3%	223.4	268.1	274.4	32.4	190
PC	219	186	84.9%	161.3	180.7	235.5	38.2	218

На графіку нижче (Рис.2) споживання обраного приладу за всі дні аналізованого періоду. Один з днів, визначений алгоритмом як найбільш репрезентативний, виділений окремим кольором для візуалізації результату.

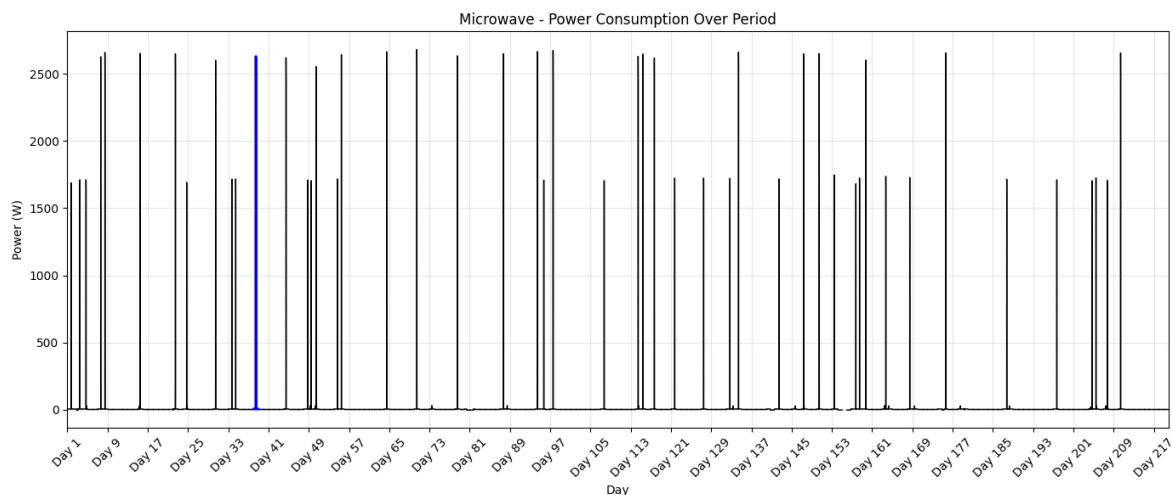


Рис.2. Репрезентативний патерн приладу

Наукова новизна та обґрунтування результатів. Запропонований підхід демонструє наукову новизну у кількох аспектах:

- Гібридна, адаптивна фільтрація: На відміну від статичних порогів, цей двоетапний підхід динамічно адаптується до розподілу даних (медіана, перцентилі), ефективно відсіюючи як

неактивні дні, так і дні з лише короткочасними піками. Це забезпечує більш релевантний набір "активних" днів для подальшого аналізу.

- Композитна метрика подібності з фіксованими вагами: Об'єднання трьох ключових аспектів (форми, амплітуди та часової синхронізації активних періодів) у єдину метрику є прагматичним рішенням. Ваги обґрунтовані інтуїтивним розумінням важливості цих аспектів для репрезентативності патерну споживання побутових пристроїв, дозволяючи уникнути надмірної чутливості до окремих характеристик.

- Прямий вибір методу без кластеризації: Замість повноцінної кластеризації з попередньо визначеною кількістю кластерів, підхід зосереджується на виборі єдиного "найбільш центрального" патерну. Це спрощує обчислення та інтерпретацію, роблячи його придатним для швидкого виявлення домінантного патерну без необхідності визначення оптимальної кількості кластерів. Отримані результати, хоча й обмежуються вибором єдиного патерну, забезпечують швидке та обґрунтоване визначення найбільш характерного профілю споживання, що є цінним для таких застосувань, як створення синтетичних профілів, базової лінії для аномалій або початкового шаблону для моделювання.

Висновки та перспективи подальших досліджень. Запропонована методологія являє собою прагматичний та ефективний підхід для ідентифікації єдиного найбільш репрезентативного добового патерну споживання енергії. Її сильні сторони включають:

- Розумна фільтрація: Двоетапний адаптивний фільтр на основі пікової та середньої потужності дозволяє ефективно відсіювати нерелевантні дні, зосереджуючись на днях з реальним, значущим споживанням.

- Багатофакторна подібність: Композитна метрика, що враховує форму, рівень потужності та синхронізацію активних періодів, забезпечує більш комплексний та інтуїтивно зрозумілий критерій "типовості".

- Простота та швидкість: Порівняно зі складними кластерними алгоритмами, вибір єдиного методу є обчислювально ефективним та легким для інтерпретації.

- Однак, існують і перспективи для подальших досліджень та вдосконалення:

- Інтеграція "Еластичних" Метрик Подібності: Замість або на додаток до L1-подібності для форми, слід розглянути застосування Dynamic Time Warping (DTW) [6] або Move Split Merge (MSM) [7]. Ці метрики краще справляються з тимчасовими зсувами та деформаціями в часових рядах.

- Формалізована Кластеризація: Замість вибору лише одного репрезентативного патерну, слід розглянути застосування K-Medoids [1] або Fuzzy C-Means [2] для виявлення кількох кластерів. Це дозволить ідентифікувати різні типи репрезентативних патернів (наприклад, "робочі дні", "вихідні").

- Розширене Вилучення Ознак: Дослідити можливість включення більш складних ознак, отриманих за допомогою вейвлет-аналізу (DWT) [4] для кількісної оцінки мінливості на різних часових масштабах, або функціонального аналізу даних.

- Адаптивне Визначення Ваг: Дослідити методи машинного навчання для динамічного визначення оптимальних ваг компонентів метрики подібності, залежно від типу даних або пристрою.

- Розвиток Системи Виявлення Аномалій: Перехід від фіксованих або простих статистичних порогів до більш складних методів виявлення аномалій, таких як ізоляційний ліс (Isolation Forest) або однокласовий SVM (One-Class SVM) [8], дозволить більш точно фільтрувати нерелевантні дні та покращити якість вхідних даних для аналізу патернів.

Подальші дослідження у цих напрямках значно підвищать точність, гнучкість та інтерпретованість систем аналізу патернів споживання енергії, сприяючи створенню більш інтелектуальних та ефективних рішень в енергетичній галузі.

Список бібліографічного опису

1. Pam, G. (1995). *K-Medoids Clustering*. Y: Encyclopedia of Database Systems. Springer, Boston, MA.
2. Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press.
3. Reynolds, D., & Roberts, D. (2005). *Gaussian Mixture Models*. Y: Encyclopedia of Speech and Language Technology. Springer, Boston, MA.
4. Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics.
5. Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series

analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971), 903–995.

6. Berndt, D. J., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In: *KDD workshop*, 10, 359-370.

7. Stefan, T., Leite, A., & Viterbo, J. (2013). Move-split-merge metric for time series. In: *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 111-120. Springer, Berlin, Heidelberg.

8. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 1-58.

9. ETH Zurich. "ECO data set (Electricity Consumption & Occupancy)".

URL: <https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption> (дата звернення: 29.05.2025).

References

1. Pam, G. (1995). *K-Medoids Clustering*. In: Encyclopedia of Database Systems. Springer, Boston, MA.

2. Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press.

3. Reynolds, D., & Roberts, D. (2005). *Gaussian Mixture Models*. In: Encyclopedia of Speech and Language Technology. Springer, Boston, MA.

4. Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics.

5. Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971), 903–995.

6. Berndt, D. J., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In: *KDD workshop*, 10, 359-370.

7. Stefan, T., Leite, A., & Viterbo, J. (2013). Move-split-merge metric for time series. In: *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 111-120. Springer, Berlin, Heidelberg.

8. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 1-58.

9. ETH Zurich. "ECO data set (Electricity Consumption & Occupancy)".

URL: <https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption> (access date: 29.05.2025).