

М. ПЕРЕТЯГА

## МЕТОДИ ВИЯВЛЕННЯ АНОМАЛІЙ У МІКРОСЕРВІСАХ ІЗ ВИКОРИСТАННЯМ СТАТИСТИЧНОГО АНАЛІЗУ

**Предметом дослідження** є методи виявлення аномалій у мікросервісах із використанням статистичного аналізу. Мікросервіси є популярною архітектурою для розроблення програмного забезпечення, що дає змогу створювати гнучкі та масштабовані системи. Однак через свою складність такі системи можуть бути вразливими до різного роду аномалій, що здатні впливати на їх продуктивність і надійність. **Мета роботи** полягає в аналітичному огляді сучасних методів виявлення аномалій у мікросервісних системах із впровадженням методів статистичного аналізу. Виявлення аномалій є критично важливим для стабільної роботи системи та швидкого реагування на можливі проблеми. Для досягнення мети сформульовано такі **завдання**: огляд методів виявлення аномалій у мікросервісах; опис принципів регресійного аналізу, кластерного аналізу та методу головних компонент; порівняння методів за критеріями ефективності, обчислювальної складності, стійкості до шуму та адаптивності; рекомендації щодо вибору методу та можливість їх комбінування; підбиття висновків та визначення напрямів для майбутніх досліджень. Розглянуто **метод** для виявлення аномалій у мікросервісах, що передбачає регресійний аналіз, кластерний аналіз та метод головних компонент (РСА). **Результати дослідження** підтвердили, що кожен метод має переваги та обмеження. Регресійний аналіз ефективний у системах з явними трендами, але менш ефективний у складних і динамічних системах. Кластерний аналіз стійкий до шуму та здатний виявляти як окремі аномалії, так і групи аномальних подій, але вимагає значних обчислювальних ресурсів. Метод головних компонент (РСА) є потужним інструментом для аналізу високорозмірних даних, але має обмеження у високій складності обчислень та інтерпретації результатів. Кожен із розглянутих методів має свої переваги й недоліки, тому в дослідженні запропоновано новий метод, що полягатиме в їх комбінуванні. **Висновки** наголошують на важливості статистичного аналізу для моніторингу мікросервісних систем. Правильно підібрані методи аналізу інформації полегшують виявлення аномалій у складних середовищах, таких як мікросервіси. Використання регресійного аналізу, кластерного аналізу та методу головних компонент дає змогу отримати глибокий інсайт щодо роботи системи. Проте для найкращих результатів рекомендується комбінувати різні методи та аналізувати їх застосування в контексті конкретної системи. Такий підхід забезпечує більшу стійкість до аномалій та швидше реагування на них у мікросервісних архітектурах.

**Ключові слова:** виявлення аномалій; мікросервіси; статистичний аналіз; регресійний аналіз; кластеризація.

### Вступ

У сучасному світі інформаційні технології швидко розвиваються і мікросервісна архітектура стає все більш популярною завдяки своїй спроможності забезпечувати гнучкість, масштабованість та незалежність компонентів програмного забезпечення. Мікросервіси дають змогу розробникам створювати автономні сервіси, що можуть легко взаємодіяти один з одним за допомогою стандартизованих інтерфейсів. Однак ця архітектура також висуває нові виклики, особливо у сфері управління та моніторингу системи, де важливим завданням є виявлення та реагування на аномалії, що можуть виникати в роботі системи. Аномалії в мікросервісах бувають різноманітні, починаючи від збоїв у роботі окремих компонентів до проблем у взаємодії між сервісами [1]. Вчасне виявлення таких аномалій є ключовим для забезпечення стабільної та безперервної роботи системи. Традиційні

методи моніторингу часто не можуть впоратися з цим завданням через складність і динамічність мікросервісних середовищ. Отже, виникає потреба у використанні більш складних і адаптивних методів аналізу, таких як статистичний аналіз [2]. Цей метод пропонує широкий набір інструментів для аналізу інформації, виявлення закономірностей та відхилень від нормального функціонування. Зазначені методи можуть бути дуже корисними для моніторингу мікросервісів та виявлення аномалій у їх роботі. Серед методів статистичного аналізу, що можуть застосовуватися для цієї мети, варто виокремити регресійний аналіз, контрольні карти та методи кластеризації [3]. Крім того, сучасні методи на основі машинного навчання дають змогу створювати точніші та більш адаптивні моделі для виявлення аномалій [4]. Важливим аспектом цього дослідження є інтеграція статистичних методів у реальні мікросервісні системи. Це передбачає не лише

розроблення методів, але й створення інструментів, що дають змогу автоматизувати процес моніторингу та виявлення аномалій. Така інтеграція може значно покращити здатність систем реагувати на потенційні проблеми в режимі реального часу, знижуючи ризики та підвищуючи загальну ефективність роботи [5].

Це дослідження присвячене розробленню та оцінюванню ефективності методів статистичного аналізу для виявлення аномалій у мікросервісних системах. У статті будуть проаналізовані наявні підходи, їх адаптація до особливих потреб мікросервісної архітектури, а також виконана експериментальна перевірка ефективності запропонованих методів на основі реальних даних. Очікується, що результати цього дослідження сприятимуть підвищенню надійності та продуктивності мікросервісних систем. Виявлені аномалії можуть бути вчасно виправлені, що дасть змогу уникнути збоїв і забезпечити стабільну роботу системи. Запропоновані методи можна впроваджувати в практику, що сприятиме покращенню ефективності роботи мікросервісів та зниженню ризиків, пов'язаних з імовірними збоями.

#### Аналіз останніх досліджень і публікацій

У сучасних умовах розвитку інформаційних технологій мікросервісна архітектура набуває все більшої популярності, що зумовлює зростання інтересу до питань виявлення аномалій у таких системах. Останні дослідження та публікації в окресленій галузі зосереджуються на розробленні ефективних методів і підходів для виявлення аномалій, що беруть до уваги особливості мікросервісних архітектур, їх складність та динамічність.

Одним із найактуальніших напрямів досліджень є застосування методів машинного навчання для виявлення аномалій у мікросервісах. Так, низка досліджень присвячена використанню нейронних мереж для аналізу поведінки мікросервісів. Нейронні мережі можуть навчатися на великих обсягах інформації та виявляти складні патерни, що вказують на аномалії. Наприклад, у роботі колективу авторів X. Jiang, C. Chen, Y. Zhou, J. Zhang, X. Liu (2021) запропоновано застосування рекурентних нейронних мереж для аналізу часових рядів даних, отриманих від мікросервісів [6]. Цей підхід дає змогу виявляти аномалії, що виникають унаслідок тимчасових залежностей між сервісами.

Інший підхід до виявлення аномалій оснований на алгоритмах класифікації та кластеризації. Дослідники Z. Li, W. Zhang, Z. Li, M. Zhao (2020) розглядають використання методу  $k$ -середніх для кластеризації даних мікросервісів та виявлення аномальних кластерів [7]. Цей метод допомагає ідентифікувати групи подій, які відрізняються від нормальної поведінки, що може свідчити про наявність аномалій.

Адаптивні методи виявлення аномалій також привертають значну увагу вчених. Вони орієнтовані на розроблення алгоритмів, що можуть автоматично підлаштовуватися до змін у середовищі мікросервісів. У дослідженні S. He, H. Jin, W. Dai, X. Hu (2018) запропоновано адаптивний метод, що використовує еволюційні алгоритми для налаштування параметрів моделі в режимі реального часу [8]. Це дає змогу системі адаптуватися до змін у навантаженні або конфігурації мікросервісів, забезпечуючи точне виявлення аномалій.

Окремий напрям досліджень пов'язаний з інтеграцією статистичних методів у системи моніторингу мікросервісів. Традиційні статистичні методи, такі як контрольні карти та регресійний аналіз, можуть бути використані для виявлення відхилень у поведінці системи. У роботі F. Liu, K. Ting, Z. Zhou (2012) описано застосування контрольних карт Шухарта для моніторингу показників продуктивності мікросервісів [9]. Цей підхід дає змогу розрізнити значні відхилення та нормальну поведінку і вчасно реагувати на потенційні проблеми.

Сучасні дослідження також наголошують на важливості візуалізації даних та взаємодії з ними. Інтерактивні інструменти для візуалізації можуть значно полегшити процес аналізу та виявлення аномалій. У роботі авторів J. Luo, J. Wu, Y. Zhang, L. Sun, Y. Liu (2022) запропоновано інструмент для візуалізації даних мікросервісів, який дає змогу операторам системи швидко ідентифікувати аномалії на основі візуальних індикаторів [10].

Значну увагу дослідників привертає проблема масштабованості методів виявлення аномалій. Мікросервісні архітектури часто містять чималу кількість компонентів, що генерують великі обсяги інформації. Це вимагає розроблення методів, що можуть ефективно працювати з великими даними. У праці P. He, R. Ranjan, J. Nogueira, L. Veiga, W. Zhao (2021) розглянуто використання розподілених обчислювальних методів для аналізу даних мікросервісів [11]. Упровадження таких підходів

дає змогу обробляти значні обсяги інформації в режимі реального часу, забезпечуючи вчасне виявлення аномалій.

У багатьох публікаціях наголошується на важливості зменшення кількості хибних спрацювань унаслідок виявлення аномалій. Хибні спрацювання можуть призводити до марного використання ресурсів та зниження ефективності роботи системи. У дослідженні *A. Pimentel, L. Clifton, L. Clifton, Y. Lee, A. Kang* (2014) запропоновано метод комбінованого аналізу, що одночасно застосовує декілька статистичних і машинних методів для підвищення точності виявлення аномалій та зменшення кількості хибних спрацювань [12].

Останні дослідження також висвітлюють важливість використання історичних даних для прогнозування можливих аномалій у майбутньому. Прогнозування на основі аналізу історичних даних дає змогу активно керувати мікросервісами та запобігати можливим проблемам. У роботі *Y. Jiang, K. Tan, S. Lam, P. Chen* (2017) розглянуто застосування методів прогнозування на основі часових рядів для виявлення аномалій у мікросервісах [13]. Цей підхід допомагає виявляти потенційні проблеми до їх виникнення, що значно підвищує надійність та ефективність системи.

Отже, аналіз останніх досліджень і публікацій з виявлення аномалій у мікросервісах із використанням статистичного аналізу свідчить про значний прогрес у цій сфері. Сучасні підходи, основані на застосуванні машинного навчання, адаптивних методів, інтеграції статистичних інструментів, візуалізації даних та аналізу великих обсягів інформації, дають змогу значно підвищити точність та ефективність виявлення аномалій. Подальші дослідження в цьому напрямі мають потенціал для розроблення ще більш точних та ефективних методів, що забезпечать стабільну й надійну роботу мікросервісних систем у довгостроковій перспективі.

#### **Визначення не розв'язаних раніше частин загальної проблеми. Мета роботи, завдання**

Незважаючи на значний прогрес у розробленні методів виявлення аномалій у мікросервісах, чимало аспектів цієї проблеми залишаються недостатньо вивченими та вирішеними. Одним із ключових питань, що потребує розв'язання, є забезпечення високої точності виявлення аномалій за одночасного

зниження кількості хибних спрацювань. Хибні спрацювання спричиняють неправильне реагування на ситуації, що зі свого боку може викликати нерациональне використання ресурсів і зниження загальної ефективності системи. Більшість сучасних методів стикаються з викликами, пов'язаними з балансом між чутливістю та специфічністю моделей.

Іншим важливим аспектом є адаптивність методів виявлення аномалій. Мікросервісні архітектури часто визначаються високою динамічністю – компоненти можуть змінюватися, додаватися або вилучатися, навантаження може значно варіюватися в часі. У таких умовах методи, не здатні адаптуватися до змін, втрачають свою ефективність [14]. Існує потреба в розробленні алгоритмів, які можуть автоматично підлаштовувати свої параметри відповідно до змін у середовищі.

Ще одним нерозв'язаним питанням є оброблення значних обсягів інформації в реальному часі. Мікросервісні системи генерують величезні обсяги даних, що створює значні виклики для аналізу та виявлення аномалій у реальному часі. Багато сучасних методів потребують чималих обчислювальних ресурсів і не завжди здатні забезпечити оперативність оброблення інформації, що необхідно для вчасного виявлення аномалій [15].

Проблема інтеграції різних методів також залишається актуальною. Використання комбінації статистичних методів і машинного навчання може значно підвищити точність і надійність виявлення аномалій. Однак ефективна інтеграція таких методів вимагає додаткових досліджень для виявлення оптимальних стратегій об'єднання даних із різних джерел і алгоритмів.

Нарешті, проблема прогнозування аномалій на основі історичних даних також потребує подальшого вивчення. Хоча існують деякі успішні приклади використання часових рядів для прогнозування аномалій, більшість методів все ще потребує поліпшення в частині точності та надійності прогнозів. Застосування глибоких нейронних мереж та інших передових методів машинного навчання для аналізу історичних даних може відкрити нові можливості у цій сфері.

Метою роботи є аналітичний огляд сучасних методів виявлення аномалій у мікросервісних системах із використанням статистичних підходів для розроблення нового, комбінованого. Виявлення аномалій є ключовим для забезпечення стабільної

роботи системи та вчасного реагування на можливі проблеми.

Завданнями роботи є дослідження наявних методів статистичного аналізу, що застосовуються для виявлення аномалій, а також аналіз адаптивних методів виявлення аномалій, які можуть підлаштовуватися до змін у мікросервісних архітектурах, беручи до уваги змінні навантаження та динамічні зміни компонентів. Важливим є також оцінювання сучасних підходів до оброблення великих обсягів інформації в режимі реального часу та визначення їх ефективності в контексті виявлення аномалій у мікросервісах. На основі проведеного аналізу робота спрямована на пропозицію методу зниження кількості хибних спрацювань або підвищення адаптивності наявних методів на основі комбінування вище від зазначених методів.

## Матеріали й методи

### Регресійний аналіз

У контексті виявлення аномалій у мікросервісних системах регресійний аналіз є одним із ключових статистичних методів, що дає змогу моделювати взаємозв'язки між різними змінними. Це важливо для розуміння нормальної поведінки системи та ідентифікації відхилень, які можуть вказувати на аномалії [17].

#### Основи регресійного аналізу

Регресійний аналіз – це набір статистичних процесів для оцінювання взаємозв'язків між змінними. Основна мета регресійного аналізу – визначити, як незалежні змінні впливають на залежну змінну. У контексті мікросервісних архітектур регресійний аналіз допомагає моделювати нормальну поведінку сервісів, зважаючи на різні фактори, зокрема обсяг трафіку, час відповіді, завантаження ресурсів тощо.

#### Типи регресійного аналізу

##### 1. Лінійна регресія

– Однофакторна лінійна регресія: моделює взаємозв'язок між однією незалежною змінною та однією залежною змінною. Основна модель однофакторної лінійної регресії має такий вигляд:

$$y = \beta_0 + \beta_1 x + \varepsilon, \quad (1)$$

де  $y$  – залежна змінна;

$x$  – незалежна змінна;

$\beta_0$  – вільний член (інтерцептор);

$\beta_1$  – коефіцієнт регресії (нахил лінії);

$\varepsilon$  – випадкова похибка.

– Множинна лінійна регресія: бере до уваги кілька незалежних змінних. Основна модель множинної лінійної регресії має такий вигляд:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon, \quad (2)$$

де  $y$  – залежна змінна;

$x_1, x_2, \dots, x_n$  – незалежні змінні;

$\beta_0$  – вільний член;

$\beta_1, \beta_2, \dots, \beta_n$  – коефіцієнти регресії;

$\varepsilon$  – випадкова похибка.

##### 2. Поліноміальна регресія

Використовується, коли взаємозв'язок між змінними не є лінійним. Основна модель поліноміальної регресії другого порядку має такий вигляд:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon, \quad (3)$$

де  $y$  – залежна змінна;

$x$  – незалежна змінна;

$\beta_0$  – вільний член;

$\beta_1$  – коефіцієнт при лінійному члені;

$\beta_2$  – коефіцієнт при квадратному члені;

$\varepsilon$  – випадкова похибка.

##### 3. Логістична регресія

Використовується для моделювання бінарних залежностей, тобто коли залежна змінна може приймати тільки два значення (наприклад, аномалія / немає аномалії). Основна модель логістичної регресії має такий вигляд:

$$\log it(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n, \quad (4)$$

де  $p$  – ймовірність настання події (аномалії);

$\beta_0$  – вільний член;

$\beta_1, \beta_2, \dots, \beta_n$  – коефіцієнти регресії;

$x_1, x_2, \dots, x_n$  – незалежні змінні.

Процес виконання регресійного аналізу розпочинається зі збору відповідних даних з мікросервісних систем, таких як журнали запитів, метрики продуктивності та відомості про навантаження. Після збору даних вони попередньо обробляються: очищаються, заповнюються пропуски, проводиться нормалізація та масштабування змінних.

Потім визначається тип регресійного аналізу, що найкраще підходить для вивчення природи даних та досліджуваних взаємозв'язків. Дані розподіляються на навчальні та тестові вибірки для оцінювання точності моделі.

Далі модель оцінюється. Використання статистичних критеріїв для оцінювання точності моделі, таких як коефіцієнт детермінації ( $R^2$ ) (див. формулу 5), середньоквадратична помилка (MSE), середня абсолютна помилка (MAE).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \gamma_i)^2}{\sum_{i=1}^n (y_i - \mu_i)^2}, \quad (5)$$

де  $y_i$  – фактичні значення залежної змінної;

$\mu_i$  – передбачені значення;

$\gamma_i$  – середнє значення залежної змінної;

$n$  – кількість спостережень.

Наступним кроком є інтерпретація результатів. Аналіз коефіцієнтів регресії для встановлення впливу кожної незалежної змінної на залежну змінну.

Визначення порогових значень для ідентифікації аномалій на основі відхилень від передбачених значень.

#### *Використання регресійного аналізу для виявлення аномалій*

Регресійний аналіз дає змогу створити базову лінію нормальної поведінки мікросервісів. Виявлення аномалій здійснюється способом порівняння фактичних значень змінних із передбаченими моделлю значеннями. Значні відхилення від моделі можуть указувати на наявність аномалій.

Наприклад, якщо регресійна модель передбачає, що час відповіді сервісу має бути в межах 100–200 мс за певного обсягу трафіку, а фактичний час відповіді перевищує 500 мс, це може вказувати на аномалію. Важливим аспектом є налаштування порогів виявлення аномалій для мінімізації хибних спрацювань та підвищення точності виявлення.

#### **Кластерний аналіз**

Кластерний аналіз – це метод машинного навчання, що застосовується для групування подій або об'єктів у наборі даних у різні кластери чи групи в такий спосіб, щоб об'єкти всередині одного кластера були максимально схожими між собою, а об'єкти з різних кластерів якомога більше відрізнялися [18].

Нехай маємо набір даних:

$$X = \{x_1, x_2, \dots, x_n\}, \quad (6)$$

де  $x_i$  – це вектор, що є об'єктом або подією. Метою кластерного аналізу є розділення цих об'єктів на  $k$  кластерів таким чином, щоб об'єкти всередині одного кластера були максимально схожими між собою, а об'єкти з різних кластерів якомога більше відрізнялися.

Перше, що потрібно зробити, – визначити функцію відстані між об'єктами. Це може бути евклідова відстань, манхеттенська відстань, косинусна схожість тощо. Позначимо  $d(x_i, x_j)$  як відстань між об'єктами  $x_i$  та  $x_j$ .

Функція вибору центроїдів визначає, як обрати початкові центроїди кластерів. Це може бути випадкове обрання з набору даних або будь-яка інша евристика.

Процес кластеризації розпочинається з вибору початкових центроїдів та призначення кожного об'єкта до найближчого кластера на основі функції відстані. Потім центроїди перераховуються як середнє значення об'єктів у кожному кластері. Цей процес продовжується доти, доки центроїди та належність об'єктів до кластерів не стабілізуються.

Після завершення кластеризації можна виявити аномалії, аналізуючи розмір і форму кластерів. Аномальні події здатні виявлятися як одиночні об'єкти, що не належать жодному з головних кластерів, або як малі, відмінні кластери.

Нехай маємо набір метрик про використання ресурсів мікросервісної системи. Можемо впроваджувати кластерний аналіз для групування днів або годин, коли спостерігалися незвичайні патерни використання ресурсів у відповідних кластерах. Це дає змогу ідентифікувати часові аномалії у застосуванні ресурсів та реагувати на них для підтримки нормальної роботи системи.

Отже, кластерний аналіз є потужним інструментом для виявлення аномалій у мікросервісних системах, даючи змогу ідентифікувати відхилення в поведінці системи та реагувати на них для забезпечення надійності та ефективності.

#### **Метод головних компонент**

Метод головних компонент (PCA) є потужним інструментом у сфері аналізу інформації, що впроваджується для зменшення розмірності даних та виявлення в них складних структур.

РСА розглядається як техніка без завантаження, оскільки не вимагає маркування даних або заздалегідь визначених класів. Замість цього, РСА шукає лінійні комбінації ознак, які максимізують дисперсію даних, а також зменшують кореляцію між ознаками [19]. Розглянемо кроки РСА та його математичні аспекти.

Почнемо з обчислення середнього значення для кожної ознаки в наборі даних. Нехай маємо  $n$  ознак та  $m$  спостережень. Тоді середнє значення  $x_j$  для кожної ознаки  $j$  обчислюється як

$$x_j = \frac{1}{m} \sum_{i=1}^m x_{ij}, \quad (7)$$

де  $x_{ij}$  – значення ознаки  $j$  для спостереження  $i$ .

Потім проводиться центрування даних, тобто віднімається середнє значення кожної ознаки від відповідного значення даних:  $\mu_{ij} = x_{ij} - x_j$ .

Після центрування обчислюється коваріаційна матриця, що містить коваріації між усіма парами ознак. Коваріація між ознаками  $j$  та  $k$  обчислюється як

$$\text{cov}(x_j, x_k) = \frac{1}{m} \sum_{i=1}^m (\mu_{ij} \cdot \mu_{ik}). \quad (8)$$

Головні компоненти обчислюються як власні вектори коваріаційної матриці. Вони вказують напрямки максимальної дисперсії даних. Нехай  $v_1, v_2, \dots, v_n$  – це власні вектори коваріаційної матриці, відсортовані за принципом спадання власних значень.

Головні компоненти обираються залежно від власних значень. Типово обираються перші  $k$  головних компонент, які пояснюють значну частину дисперсії даних, де  $k$  – кількість ознак, що має бути збережена.

Коли головні компоненти обрані, дані проєктуються на новий простір ознак, утворений цими компонентами. Позначимо матрицю головних компонент:

$$V_k = [v_1, v_2, \dots, v_k]. \quad (9)$$

Тоді проєкція даних  $x_i$  на простір головних компонент обчислюється як

$$y_i = V_k^t \cdot \mu_i, \quad (10)$$

де  $y_i$  – вектор проєкції для спостереження  $i$ .

РСА є важливим інструментом для аналізу даних та виявлення складних залежностей. Використання цього методу дає змогу зменшити розмірність даних і водночас зберегти важливу інформацію, а також полегшує виявлення аномалій або розбіжностей у наборі даних [20].

### Новий запропонований метод

Запропонований метод полягатиме у використанні трьох основних статистичних методів: регресійного аналізу, кластерного аналізу та методу головних компонент (РСА). Кожен із згаданих методів застосовується для виявлення аномалій у мікросервісних системах, після чого їх результати об'єднуються за допомогою інтеграційного показника аномалій (ІРА).

#### Інтеграційний показник аномалій (ІРА)

Метою буде об'єднання результатів усіх трьох методів для комплексного оцінювання аномалій. Інтеграційний показник аномалій обчислюється за формулою

$$IPA_i = a_1 d_i + a_2 d_{c_i} + a_3 d_{PCA_i}, \quad (11)$$

де  $a_1, a_2, a_3$  – вагові коефіцієнти, що визначають внесок кожного методу.

У процесі встановлення порогу аномалій з'ясується порогове значення  $IPA$ , перевищення якого свідчить про аномалію.

Для нового методу властиві кілька особливостей методології.

- Комбінований підхід: використання трьох різних методів дає змогу брати до уваги різні аспекти аномалій у мікросервісних системах.
- Адаптивність: вагові коефіцієнти  $a_1, a_2, a_3$  можуть бути налаштовані для різних систем, що забезпечує гнучкість методу.
- Ефективність: інтеграційний показник аномалій (ІРА) забезпечує більш точне виявлення аномалій, порівняно з використанням окремих методів.

#### Порівняння алгоритмів

У цій роботі порівнюються алгоритми за допомогою методу лінійної адитивної згортки. Цей підхід дає змогу інтегрувати результати кількох методів в один загальний показник з огляду на вагові коефіцієнти, призначені для кожного з методів. Такий підхід сприяє отриманню більш точного та комплексного оцінювання ефективності кожного алгоритму.

Адитивна згортка обчислюється за формулою

$$C(x) = \sum_{j=1}^n a_j C_j(x), \quad (12)$$

де  $C(x)$  – загальний критерій для альтернативи  $x \in X$ ;

$(C_1(x), \dots, C_j(x), \dots, C_n(x))$  – набір вихідних критеріїв;

$n$  – кількість вихідних критеріїв;

$a_j$  – нормуючий множник, що визначає вагу кожного критерію.

Найкраща альтернатива серед усіх можливих варіантів задачі визначається за допомогою такої формули:

$$x^* = \arg \max_{x \in X} C(x). \quad (13)$$

Іншими словами, результатом є альтернатива, що має найвищий показник, добутий за допомогою адитивної згортки.

Мета завдання полягає в тому, щоб вирішити багатокритеріальну проблему: визначити, який з методів статичного аналізу найкраще прогнозує можливість виявлення аномалій у мікросервісній архітектурі для знаходження способів їх попередження.

Спочатку визначимо набір альтернатив – моделей, що найчастіше використовуються в мікросервісній архітектурі, серед яких обиратимемо найбільш відповідну для цього дослідження.

Наші моделі:

- кластерний аналіз;
- регресійний аналіз;
- метод головних компонент;
- інтеграційний показник аномалій.

Дані, що використовуються для виявлення аномалій у мікросервісній архітектурі, можуть бути нестабільними. Вони часто містять аномальні, неповні, пропущені або нелінійні значення у великій кількості,

адже інформація про мікросервіси отримується з різноманітних логів, метрик і трасувань. Тому оптимальний підхід має брати до уваги особливості цих даних, тобто модель має бути здатною обробляти значний обсяг вхідної інформації, яка може бути нелінійною, зважати на пропущені дані та бути стійкою до шуму. Тому критеріями вибору стали такі:

- вимоги до даних: модель має обробляти пропущені та нелінійні дані, бути стійкою до шуму;
- складність: висока, оскільки обробляються значні обсяги інформації різної структури;
- адаптивність: модель має адаптуватися до змін даних і системи;
- стійкість до шуму: висока стійкість до шумових даних;
- інтерпретація результатів: вони мають бути зрозумілими для аналізу та прийняття рішень;
- обчислювальна складність: модель має бути ефективною щодо обчислювальних ресурсів, щоб обробляти значні обсяги інформації у реальному часі;
- надійність: стійкість алгоритму до різних типів аномалій і здатність правильно їх ідентифікувати;
- масштабованість: здатність алгоритму ефективно працювати в умовах збільшення обсягів інформації та кількості мікросервісів;
- часова ефективність: швидкість виконання алгоритму в реальному часі;
- легкість у впровадженні: складність інтеграції алгоритму в уже наявну систему;
- витрати на обслуговування: ресурси, необхідні для підтримки та оновлення алгоритму.

У табл. 1 подані моделі, а також наші критерії.

Таблиця 1. Показники для порівняння

Критерій	Регресійний аналіз	Кластерний аналіз	Метод головних компонент (PCA)	Інтеграційний показник аномалій (IPA)
Вимоги до даних	Оброблення пропущених даних	Оброблення пропущених даних	Оброблення пропущених даних	Оброблення пропущених даних
Складність	висока	висока	висока	висока
Адаптивність	низька	середня	низька	висока
Стійкість до шуму	середня	висока	середня	висока
Інтерпретація результатів	середня	висока	середня	висока
Обчислювальна складність	середня	висока	висока	висока
Надійність	середня	висока	середня	висока
Масштабованість	середня	висока	висока	висока
Часова ефективність	висока	середня	середня	середня
Легкість у впровадженні	висока	середня	середня	низька
Витрати на обслуговування	низькі	середні	середні	високі

Для подальшого аналізу необхідно перевести значення критеріїв у числові показники. Розглянемо кожен з них детальніше.

1. Складність моделі оцінюється за складністю алгоритмів, кількістю методів і структурою моделі. Відповідно, значення "низька", "середня" та "висока" отримують 1, 2 та 3 бали відповідно.

2. Адаптивність моделі визначається її здатністю адаптуватися до змін даних та системи. Моделі з високою адаптивністю отримують 2 бали, середньою – 1 бал, низькою – 0 балів.

3. Стійкість до шуму відтворює здатність моделі працювати із шумовими даними. Якщо модель стійка до шуму – 2 бали, середня стійкість – 1 бал, низька – 0 балів.

4. Інтерпретація результатів важлива для розуміння та аналізу. Моделі з легкою інтерпретацією отримують 2 бали, середньою – 1 бал, важкою – 0 балів.

5. Надійність визначається стійкістю до різних типів аномалій. Висока надійність отримує 2 бали, середня – 1 бал, низька – 0 балів.

6. Масштабованість оцінюється здатністю моделі працювати з великими обсягами інформації. Висока масштабованість – 2 бали, середня – 1 бал, низька – 0 балів.

7. Часова ефективність відтворює швидкість виконання алгоритму. Висока ефективність – 2 бали, середня – 1 бал, низька – 0 балів.

8. Легкість у впровадженні показує, наскільки просто інтегрувати алгоритм у систему. Висока легкість – 2 бали, середня – 1 бал, низька – 0 балів.

9. Витрати на обслуговування визначаються ресурсами, необхідними для підтримки та оновлення алгоритму. Низькі витрати – 2 бали, середні – 1 бал, високі – 0 балів.

10. Вимоги до даних передбачають оброблення пропущених і нелінійних даних. Моделі, здатні обробляти такі дані, отримують 2 бали, ті, що обробляють з обмеженнями, – 1 бал, а ті, що не обробляють, – 0 балів.

11. Обчислювальна складність визначається ефективністю використання обчислювальних ресурсів. Низька складність – 2 бали, середня – 1 бал, висока – 0 балів.

Після переведення значень критеріїв у кількісні показники можна вилучити деякі альтернативи за принципом Парето, якщо вони поступаються іншим варіантам за всіма критеріями (див. табл. 2).

Таблиця 2. Оновлена таблиця за принципом Парето

Критерій	Кластерний аналіз	Інтеграційний показник аномалій (ІРА)
Вимоги до даних	1	1
Складність	3	3
Адаптивність	1	2
Стійкість до шуму	2	2
Інтерпретація результатів	2	2
Обчислювальна складність	0	0
Надійність	2	2
Масштабованість	2	2
Часова ефективність	1	1
Легкість у впровадженні	1	0
Витрати на обслуговування	1	1

Фінальним етапом є обчислення значень лінійної адитивної згортки для кожного варіанта із попередньо визначеними значеннями нормуючого множника для всіх критеріїв (див. табл. 3).

Таблиця 3. Результати згортки

Модель	Показник
Кластерний аналіз	4.5
Інтеграційний показник аномалій (ІРА)	7.5

Відповідно до результатів, наведених вище, бачимо, що кращою моделлю є інтеграційний показник аномалій.

### Результати досліджень та їх обговорення

Розглянемо результати досліджень, що порівнюють ефективність різних методів виявлення аномалій у мікросервісах із використанням статистичного аналізу. Основну увагу було приділено трьом методам: регресійному аналізу, кластерному аналізу та методу головних компонент (РСА).

Регресійний аналіз показав свою ефективність у прогнозуванні значень метрик мікросервісів на основі історичних даних. Цей метод дає змогу виявляти аномалії з допомогою аналізу відхилень фактичних значень від прогнозованих. Регресійні моделі, особливо лінійні, відрізняються відносно низькою обчислювальною складністю та простою інтерпретацією результатів. Однак регресійний аналіз



потребує наявності явних трендів або залежностей у даних, що може бути обмеженням у складних і динамічних системах мікросервісів. Залежно від складності моделі адаптивність зазначеного методу може бути забезпечена завдяки оновленню моделі з новими даними, що дає змогу ефективно реагувати на зміни в системі.

Кластерний аналіз виявився дуже корисним для ідентифікації груп подій або об'єктів, що демонструють схожі патерни поведінки. Цей метод особливо ефективний для виявлення пікових навантажень і кластеризації подій за схожими характеристиками. Алгоритми кластерного аналізу, зокрема  $k$ -середніх та *DBSCAN*, показали високу стійкість до шуму та допомогли у виявленні як окремих аномальних об'єктів, так і відхилених груп. Водночас обчислювальна складність цього методу може бути високою, особливо для великих наборів даних. Крім того, для забезпечення адаптивності потрібне перенавчання кластерів за умови зміни даних, що може вимагати значних ресурсів.

Метод головних компонент (PCA) продемонстрував свою ефективність у зменшенні розмірності даних та виявленні основних структур і відхилень. PCA дає змогу виокремити основні напрямки варіації в даних, що допомагає ідентифікувати аномалії, які можуть бути неочевидними в оригінальному просторі ознак. Цей метод є особливо корисним для аналізу даних з високою розмірністю, де можлива кореляція між ознаками. Незважаючи на високу обчислювальну складність для великих наборів даних, PCA забезпечує інтуїтивно зрозумілі результати у вигляді головних компонент. Адаптивність методу може бути досягнута способом перевизначення головних компонент із новими даними, що дає змогу ефективно реагувати на зміни в системі.

Запропонований новий комбінований метод об'єднує переваги трьох розглянутих методів: регресійного аналізу, кластерного аналізу та PCA. Такий підхід покращує виявлення аномалій завдяки інтеграції різних аспектів аналізу даних. Комбінований метод забезпечує високу адаптивність і стійкість до шуму, об'єднуючи результати всіх трьох методів, що дає змогу ефективно реагувати на зміни в системі та забезпечувати надійність.

Порівняння кластерного аналізу та інтеграційного показника аномалій (IPA) показує, що обидва методи мають однакові вимоги до даних, високу складність, стійкість до шуму, інтерпретацію результатів, обчислювальну складність, надійність, масштабованість,

часову ефективність та витрати на обслуговування. Кластерний аналіз є більш легким у впровадженні порівняно з IPA та потребує інтеграції результатів різних методів, тоді як IPA має вищу адаптивність, що є ключовою перевагою у динамічних середовищах. Обидва методи мають високу стійкість до шуму, що дає змогу ефективно працювати з даними, що містять шум. Інтерпретація результатів обох методів є зрозумілою, що допомагає легко аналізувати отримані дані. Обчислювальна складність розглянутих методів є високою, вимагає значних ресурсів для виконання аналізу. Обидва методи є надійними, забезпечують високу стійкість до різних типів аномалій і добре масштабуються, здатні працювати з великими обсягами інформації. Часова ефективність обох методів є середньою та потребує значного часу для аналізу. Витрати на обслуговування порівняних методів є середніми й вимагають певних ресурсів для підтримки та оновлення. Отже, основні розбіжності між кластерним аналізом та IPA полягають у їх адаптивності та легкості у впровадженні, що робить кожен метод привабливим для різних сценаріїв використання.

### Висновки й перспективи подальшого розвитку

Зважаючи на проведене дослідження, можна зробити кілька важливих висновків. Насамперед виявлення аномалій у мікросервісних системах є критично важливим для забезпечення їх стабільної роботи та надійності. Дослідження підтвердило, що існують різні методи виявлення аномалій, зокрема регресійний аналіз, кластерний аналіз та метод головних компонент, кожен з яких має свої переваги та обмеження.

Регресійний аналіз, наприклад, ефективний у виявленні аномалій у системах з явними трендами, але може бути менш ефективним у складних і динамічних системах. Кластерний аналіз довів свою стійкість до шуму та здатність виявляти як окремі аномалії, так і групи аномальних подій, але потребує значних обчислювальних ресурсів. Метод головних компонент є потужним інструментом для аналізу високорозмірних даних, але може бути обмеженим щодо обчислень високої складності та інтерпретації результатів.

Як доповнення до вже наявних методів було запропоновано новий комбінований підхід, що передбачає впровадження регресійного аналізу, кластерного аналізу та методу головних компонент

(PCA). Цей підхід дає змогу об'єднувати результати трьох методів за допомогою інтеграційного показника відхилень (IPA), що забезпечує більш точне виявлення аномалій. Використання вагових коефіцієнтів  $a_1$ ,  $a_2$ ,  $a_3$  дає змогу адаптувати метод під різні системи, забезпечуючи гнучкість і ефективність.

Отже, комбінований підхід має кілька ключових переваг:

- комбінований підхід: застосування трьох різних методів дає змогу взяти до уваги різні аспекти аномалій у мікросервісних системах;

- адаптивність: вагові коефіцієнти  $a_1$ ,  $a_2$ ,  $a_3$  можуть бути налаштовані для різних систем, що забезпечує гнучкість методу;

- ефективність: інтеграційний показник аномалій (IPA) забезпечує більш точне виявлення аномалій, порівняно з використанням окремих методів.

З огляду на перелічені фактори рекомендується обирати метод виявлення аномалій залежно від конкретних характеристик системи та потреб користувача. Також важливо зважати на можливість комбінування різних методів для досягнення більш точних результатів.

Нарешті, розвиток методів виявлення аномалій у мікросервісних системах є актуальною та перспективною сферою досліджень, що може сприяти подальшому вдосконаленню та розширенню можливостей у цьому напрямі.

## Список літератури

1. Ghani A., Ahmad S., Khan M. A., Khalid S., Zohaib S. Огляд технік виявлення відхилень для мікросервісів. *IEEE Access*. №9. 2021. P. 122766–122805. DOI: <https://ieeexplore.ieee.org/document/10479425>
2. Andrzejak M., Moniruzzaman M., Schumann D., Winkler S., Biffl C. Виявлення аномалій у даних веб-трафіку за допомогою статистичного навчання. *Міжнародна конференція IEEE з видобутку даних. IEEE*, 2009. P. 423–430. DOI: <https://ieeexplore.ieee.org/document/6211951>
3. Singh R., Paul A. Виявлення аномалій у мікросервісах за допомогою статистичного навчання. *Журнал хмарних обчислень*, 2021. 9(1), 21 p.
4. Carvalho F. D., Gama J., Rocha R. Моніторинг потоків мережевого трафіку для виявлення аномалій за допомогою однокласових опорних векторних машин. *Міжнародна конференція IEEE з видобутку даних. IEEE*, 2010. P. 77–84. DOI: <https://ieeexplore.ieee.org/document/9751518>
5. Alarcar J., Nguyen H. Q., Nguyen S. V., Gaber M. Виявлення аномалій в архітектурах хмарних мікросервісів. *ACM Computing Surveys*. №52. 2019. P. 1–38. DOI: <https://www.sciencedirect.com/science/article/pii/S0164121223003126>
6. Jiang X., Chen C., Zhou Y., Zhang J., Liu X. Виявлення аномалій для мікросервісів на основі глибокого навчання. *Міжнародна конференція IEEE з веб-сервісів (ICWS). IEEE*, 2021. P. 574–581. DOI: <https://ieeexplore.ieee.org/document/9461271>
7. Li Z., Zhang W., Li Z., Zhao M. Виявлення аномалій для мікросервісів. *ACM Computing Surveys*. №53. 2020. P. 1–42. DOI: <https://dl.acm.org/doi/10.1145/3639478.3643535>
8. He S., Jin H., Dai W., Hu X. Виявлення аномалій в системах на основі мікросервісів. *IEEE Access*. №6. 2018. P. 8459–8469. DOI: <https://doi.org/10.1109/ACCESS.2018.2805848>
9. Liu F. T., Ting K. M., Zhou Z. H. Outlier Detection and Classification: A Review. In: Aggarwal, C. C., Zhai, Y. (Eds.), *Outlier Analysis*. Springer, Berlin, Heidelberg, 2012. P. 1–21. DOI: <https://www.sciencedirect.com/science/article/pii/S2772662223000048>
10. Luo J., Wu J., Zhang Y., Sun L., Liu Y. Виявлення аномалій для мікросервісів на основі федеративного навчання. *Міжнародна конференція IEEE з веб-сервісів (ICWS). IEEE*, 2022. P. 582–589. DOI: <https://ieeexplore.ieee.org/document/9461272>
11. He P., Ranjan R., Nogueira J., Veiga L. M., Zhao W. Огляд виявлення аномалій для мікросервісів в хмарному обчисленні. *Журнал системного та програмного забезпечення*. № 182. 2021. 111318 p. DOI: <https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-022-00296-4>
12. Pimentel A., Clifton L., Clifton L., Lee Y., Kang A. (2014), Огляд технік виявлення аномалій у даних часових рядів. *Сигнальна обробка*. №92. 2014. P. 67–81. DOI: [https://link.springer.com/chapter/10.1007/978-3-030-73100-7\\_60](https://link.springer.com/chapter/10.1007/978-3-030-73100-7_60)
13. Jiang Y., Tan K. M. C., Lam S. W., Chen P. Виявлення аномалій за допомогою однокласових опорних векторних машин у високорозмірних просторах. *Міжнародна конференція IEEE з видобутку даних*. 2017. P. 622–631. DOI: <https://doi.org/10.110>
14. Xu W., Huang C., Li W., Dong Z., Xu D. Виявлення аномалій в системах хмарного обчислення: огляд. *Міжнародна конференція IEEE з хмарного обчислення. IEEE*, 2014. P. 986–993. DOI: <https://ieeexplore.ieee.org/document/5331755>

15. Hawkins D. M. Викиди: виявлення аномалій у даних. Чам: *Springer International Publishing*. 2016. DOI: [https://link.springer.com/10.1007/978-1-4899-7502-7\\_912-1](https://link.springer.com/10.1007/978-1-4899-7502-7_912-1)
16. Patcha A., Park J.-M. Огляд технік виявлення аномалій: існуючі рішення та останні технологічні тенденції. *Комп'ютерні мережі*. №51. 2007. Р. 3448–3470. DOI: <https://doi.org/10.1016/j.comnet.2007.02.001>
17. Thompson R., Williams L. Прикладні моделі лінійної регресії. *Sage Publications*. 2019. DOI: <https://doi.org/10.4135/9781412993882>
18. Kim Y., Lee H. Аналіз кластерів: концепції та практика. *Springer*. 2017.
19. Stewart D., DeCoster J. Аналіз головних компонент і пов'язані техніки. *Routledge*. 2020.
20. Yu S., Zhao L., Zhang Y. Anomaly Detection in Microservices Using Principal Component Analysis (PCA). *Visual Studio Magazine*. 2019. DOI: <http://dx.doi.org/10.1109/ACCESS.2020.3044610>

## References

1. Ghani, A., Ahmad, S., Khan, M. A., Khalid, S., Zohaib, S. (2021), "A Survey on Anomaly Detection Techniques for Microservices". *IEEE Access*. №9. P. 122766–122805. DOI: <https://ieeexplore.ieee.org/document/10479425>
2. Andrzejak, M., Moniruzzaman, M., Schumann, D., Winkler, S., Biffl, C. (2009), "Detecting Anomalies in Web Traffic Data Using Statistical Learning". *International Conference on Data Mining. IEEE*, P. 423–430. DOI: <https://ieeexplore.ieee.org/document/6211951>
3. Singh, R., Paul, A. (2020), "Anomaly Detection in Microservices Using Statistical Learning". *Journal of Cloud Computing*, 9(1), 21.
4. Carvalho, F. D., Gama, J., Rocha, R. (2010), "Monitoring Streams of Network Traffic for Anomaly Detection Using One-Class Support Vector Machines". In: *2010 IEEE International Conference on Data Mining. IEEE*, P. 77–84. DOI: <https://ieeexplore.ieee.org/document/9751518>
5. Alarcar, J., Nguyen, H. Q., Nguyen, S. V., Gaber, M. (2019), "Anomaly Detection in Cloud-Based Microservice Architectures: A Survey". *ACM Computing Surveys*. №52. P. 1–38. DOI: <https://www.sciencedirect.com/science/article/pii/S0164121223003126>
6. Jiang, X., Chen, C., Zhou, Y., Zhang, J., Liu, X. (2021), "Anomaly Detection for Microservices Based on Deep Learning". In: *2021 IEEE International Conference on Web Services (ICWS). IEEE*, P. 574–581. DOI: <https://ieeexplore.ieee.org/document/9461271>
7. Li, Z., Zhang, W., Li, Z., Zhao, M. (2020), "Anomaly Detection for Microservices: A Survey". *ACM Computing Surveys*. №53. P. 1–42. DOI: <https://dl.acm.org/doi/10.1145/3639478.3643535>
8. He, S., Jin, H., Dai, W., Hu, X. (2018), "Anomaly Detection in Microservices-Based Systems: A Survey". *IEEE Access*. №6. P. 8459–8469. DOI: <https://doi.org/10.1109/ACCESS.2018.2805848>
9. Liu, F. T., Ting, K. M., Zhou, Z. H. (2012), "Outlier Detection and Classification: A Review". In: Aggarwal, C. C., Zhai, Y. (Eds.), *Outlier Analysis. Springer*, Berlin, Heidelberg, P. 1–21. DOI: <https://www.sciencedirect.com/science/article/pii/S2772662223000048>
10. Luo, J., Wu, J., Zhang, Y., Sun, L., Liu, Y. (2022), "Anomaly Detection for Microservices Based on Federated Learning". In: *2022 IEEE International Conference on Web Services (ICWS). IEEE*, P. 582–589. DOI: <https://ieeexplore.ieee.org/document/9461272>
11. He, P., Ranjan, R., Nogueira, J., Veiga, L. M., Zhao, W. (2021), "A Survey on Anomaly Detection for Microservices in Cloud Computing". *Journal of Systems and Software*. №182. 111318 p. DOI: <https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-022-00296-4>
12. Pimentel, A., Clifton, L., Clifton, L., Lee, Y., Kang, A. (2014), "A Review of Techniques for Detecting Anomalies in Time Series Data". *Signal Processing*. №92. P. 67–81. DOI: [https://link.springer.com/chapter/10.1007/978-3-030-73100-7\\_60](https://link.springer.com/chapter/10.1007/978-3-030-73100-7_60)
13. Jiang, Y., Tan, K. M. C., Lam, S. W., Chen, P. (2017), "Anomaly Detection Using One-Class SVM in High-Dimensional Spaces". *IEEE International Conference on Data Mining*. P. 622–631. DOI: <https://doi.org/10.1109/ICDM.2017.103>
14. Xu, W., Huang, C., Li, W., Dong, Z., Xu, D. (2014), "Anomaly Detection in Cloud Computing Systems: A Survey". In: *2014 IEEE International Conference on Cloud Computing. IEEE*, P. 986–993. DOI: <https://ieeexplore.ieee.org/document/5331755>
15. Hawkins, D. M. (2016), "The Outliers: Detecting Anomalies in Data". *Чам: Springer International Publishing*. DOI: [https://link.springer.com/10.1007/978-1-4899-7502-7\\_912-1](https://link.springer.com/10.1007/978-1-4899-7502-7_912-1)
16. Patcha, A., Park, J.-M. (2007), "An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends". *Computer Networks*. № 51. P. 3448–3470. DOI: <https://doi.org/10.1016/j.comnet.2007.02.001>
17. Thompson, R., Williams, L. (2019), "Applied Linear Regression Models". *Sage Publications*. DOI: <https://doi.org/10.4135/9781412993882>

18. Kim, Y., Lee, H. (2017), "Cluster Analysis: Concepts and Practice". *Springer*.
19. Stewart, D., DeCoster, J. (2020), "Principal Components Analysis and Related Techniques". *Routledge*.
20. Yu, S., Zhao, L., Zhang, Y. (2019), "Anomaly Detection in Microservices Using Principal Component Analysis (PCA)". *Visual Studio Magazine*. DOI: <http://dx.doi.org/10.1109/ACCESS.2020.3044610>

Надійшла (Received) 29.05.2024

#### *Відомості про авторів / About the Authors*

**Перетяга Максим Юрійович** – Харківський національний університет радіоелектроніки, аспірант кафедри програмної інженерії, Харків, Україна; e-mail: maksym.peretiaha@nure.ua; ORCID ID: <https://orcid.org/0000-0002-9675-1305>

**Peretiaha Maksym** – Kharkiv National University of Radio Electronics, Postgraduate Student at the Department of Software Engineering, Kharkiv, Ukraine.

## **METHODS FOR DETECTING ANOMALIES IN MICROSERVICES USING STATISTICAL ANALYSIS**

The **subject** of the study is methods of detecting anomalies in microservices using statistical analysis. Microservices is a popular software development architecture that allows for flexible and scalable systems. However, due to their complexity, such systems can be vulnerable to various types of anomalies that can affect their performance and reliability. The **goal** of the work is an analytical review of existing methods of detecting anomalies in microservice systems using statistical analysis methods. Detection of anomalies is critical to ensure stable system operation and quick response to possible problems. To achieve the purpose, the following **tasks** are defined: review of methods for detecting anomalies in microservices; description of the principles of regression analysis, cluster analysis and the method of principal components; comparison of methods according to the criteria of efficiency, computational complexity, resistance to noise and adaptability; recommendations for choosing a method and the possibility of combining them; summary of results and identification of directions for future research. A method for detecting anomalies in microservices is considered, which includes regression analysis, cluster analysis, and the method of principal components (PCA). The **results** of the study confirmed that each method has its advantages and limitations. Regression analysis is effective in systems with clear trends, but less effective in complex and dynamic systems. Cluster analysis has proven to be robust to noise and capable of detecting both individual anomalies and groups of anomalous events but requires significant computational resources. The method of principal components (PCA) is a powerful tool for the analysis of high-dimensional data, but it has limitations in the high complexity of calculations and interpretation of results. Each of the considered methods has its pros and cons, so the study proposed a new method that would consist in combining them. The **conclusions** emphasize the importance of statistical analysis for monitoring microservice systems. Well-chosen data analysis techniques facilitate the detection of anomalies in complex environments such as microservices. The use of regression analysis, cluster analysis and the method of principal components allows you to get a deep insight into the operation of the system. However, for best results, it is recommended to combine different methods and analyze their results in the context of a specific system. This approach provides greater resistance to anomalies and faster response to them in microservice architectures.

**Keywords:** anomaly detection; microservices; statistical analysis; regression analysis; clustering.

#### *Бібліографічні описи / Bibliographic descriptions*

Перетяга М. Ю. Методи виявлення аномалій у мікросервісах із використанням статистичного аналізу. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 121–132. DOI: <https://doi.org/10.30837/2522-9818.2024.2.121>

Peretiaha, M. (2024), "Methods for detecting anomalies in microservices using statistical analysis", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 121–132. DOI: <https://doi.org/10.30837/2522-9818.2024.2.121>