© Martyniuk I.V., Okhrimenko T.O. Methodology and classification of open-source ML methods for IT monitoring based on the Zabbix system // *Ukrainian Scientific Journal of Information Security, 2025, vol. 31, issue 2, pp. 87-92.* ©

DOI: 10.18372/2225-5036.31.20700

МЕТОДОЛОГІЯ ТА КЛАСИФІКАЦІЯ ML-МЕТОДІВ З ВІДКРИТИМ КОДОМ ДЛЯ IT-МОНІТОРИНГУ НА БАЗІ СИСТЕМИ ZABBIX

Ігор Мартинюк[1,2], Тетяна Охріменко[1,2]

[1]*Державний університет «Київський авіаційний інститут», Україна*
[2]*Державний науково-дослідний інститут технологій кібербезпеки та захисту інформації, Україна*

МАРТИНЮК Ігор Вадимович, аспірант
Рік та місце народження: 1988 рік, с. Верхнєдніпровський, Росія.
Освіта: Національний авіаційний університет, 2011 рік.; аспірант, Державний університет «Київський авіаційний інститут» з 2024 року.
Посада: інженер Державного науково-дослідного інституту технологій кібербезпеки та захисту інформації з 2023 року.
Наукові інтереси: кібербезпека, моніторинг.
Публікації: більше 10 наукових публікацій, серед яких наукові статті, матеріали та тези доповідей на конференціях.
E-mail: imartyniukiv@gmail.com
ORCID: 0009-0003-5565-0828

ОХРІМЕНКО Тетяна Олександрівна, к.т.н., ст. дослідник
Рік та місце народження: 1990 рік, м. Вінниця, Україна.
Освіта: Національний авіаційний університет, 2012 рік.
Посада: заступник декана з наукової роботи Факультету комп'ютерних наук та технологій Державний університет «Київський авіаційний інститут»; провідний науковий співробітник ДержНДІ.
Наукові інтереси: інформаційна безпека, реагування на інциденти інформаційної безпеки, квантова криптографія.
Публікації: більше 100 наукових публікацій, серед яких монографії, наукові статті, матеріали та тези доповідей на конференціях, патенти та авторські свідоцтва.
E-mail: t.okhrimenko@npp.kai.edu.ua
ORCID: 0000-0001-9036-6556

Анотація. *У даній статті розглянуто застосування методів машинного навчання з відкритим кодом для задач IT-моніторингу на базі системи Zabbix. Проаналізовано підходи до виявлення аномалій, прогнозування часових рядів та аналізу лог-файлів, а також їх обмеження у контексті експлуатаційного моніторингу. Запропоновано методологію інтеграції зовнішніх ML-модулів із Zabbix та класифікаційну схему використання ML-моделей залежно від типу даних і задач. Виконано порівняльний аналіз ML-підходів і сформульовано рекомендації щодо їх практичного застосування з урахуванням вимог для досягнення цільового рівня якості сервісу (SLO).*

Ключові слова: *кібербезпека, інформаційні технології, IT-моніторинг, Zabbix, машинне навчання, ML-моделі, виявлення аномалій, прогнозування, аналіз логів.*

Вступ

При зростанні масштабів і складності сучасних IT-інфраструктур системи моніторингу повинні обробляти значні обсяги метрик і даних в режимі реального часу. Характерними рисами таких інфраструктур є динамічність навантаження, складні міжкомпонентні залежності та часті зміни конфігурацій, що ускладнюють використання статичних правил контролю за станом кожного елементу системи. Класичні механізми виявлення інцидентів, що базуються на списку правил, часто не здатні адаптуватися до динамічної поведінки систем, що призводить до хибних спрацювань або пропуску критичних інцидентів [1, 2]. У зв'язку з цим у наукових і прикладних дослідженнях активно розглядаються методи машинного навчання (ML) як засіб підвищення точності виявлення аномалій, прогнозування показників і семантичного аналізу логів, що отримують системи IT-моніторингу [3].

Важливу роль відіграють рішення з відкритим кодом за рахунок своєї доступності для широкого використання, які можуть бути інтегровані в існуючі моніторингові платформи. Однією з таких платформ є Zabbix – поширена система централізованого моніторингу, що забезпечує збір метрик, логів, використання тригерів та сповіщень. З іншого боку архітектура Zabbix орієнтована більшою мірою на порогові та статистичні методи аналізу даних, що обмежує застосування сучасних підходів машинного навчання без використання зовнішніх компонентів. Це створює потребу в дослідженні методів інтеграції зовнішніх ML-моделей для виконання задач anomaly detection, forecasting та log analysis, орієнтованих на підвищення рівня якості сервісу (Service Level Objective - SLO) [4, 5].

Аналіз існуючих досліджень

У задачах IT-моніторингу аномалії, як правило, проявляються у вигляді відхилень значень метрик (CPU, використання пам'яті, затримки, кількість помилок) від типової поведінки системи. У сучасних дослідженнях виділяють три основні групи

підходів до виявлення аномалій у часових рядах: статистичні, класичні ML та deep learning методи [3].

По-перше, статистичні методи (ARIMA, EWMA, STL-декомпозиція) ґрунтуються на припущенні стаціонарності у часових рядах. Вони використовуються для формування очікуваного значення метрики з подальшим порівнянням його з фактичними спостереженнями. Такі методи відзначаються простотою реалізації та високою інтерпретованістю, однак демонструють обмежену ефективність у багатовимірних та нестаціонарних середовищах [6].

По-друге, класичні методи машинного навчання (Isolation Forest, One-Class SVM, Local Outlier Factor) не потребують розмічених вибірок і дозволяють виявляти аномалії без розмічених даних [7]. Вони широко представлені у відкритих програмних бібліотеках (зокрема PyOD) та можуть застосовуватися для аналізу даних, експортованих із системи Zabbix. Основними обмеженнями цих підходів є слабке врахування часової залежності та складність інтерпретації отриманих результатів.

По-третє, методи глибокого навчання (Autoencoder, LSTM-AE, Variational Autoencoder, Transformer) широко застосовуються для виявлення аномалій у часових рядах і демонструють високу ефективність у складних, нестаціонарних середовищах, проте потребують значних обчислювальних ресурсів та є більш складними в експлуатації [8].

Прогнозування часових рядів у системах моніторингу використовується для планування ресурсів, запобігання деградації сервісів і прогнозування порушень цільових показників якості обслуговування (SLO) [9]. У цьому напрямі виділяють такі підходи.

Класичні методи прогнозування (ARIMA, Holt–Winters, Prophet) ефективні для моделювання трендів і сезонності та можуть застосовуватися безпосередньо в рамках прогнозних тригерів Zabbix. В той же час вони мають обмежену здатність адаптуватися до раптових змін навантаження [6].

ML/DL-підходи (LSTM, GRU, Transformer) дозволяють враховувати багатовимірні залежності між метриками та прогнозувати не лише майбутні значення показників, а й імовірність порушення SLO [10]. Такі методи є перспективними для проактивного моніторингу, однак характеризуються підвищеною складністю впровадження та експлуатації.

Аналіз логів є важливим компонентом діагностики інцидентів і виявлення прихованих збоїв. У сучасних дослідженнях застосовуються наступні підходи.

Традиційні rule-based та шаблонні методи (regular expressions, Drain, Spell) використовуються для структуризації неформатованих лог-повідомлень та широко використовується в системах моніторингу [11]. Вони ефективні на етапі попередньої обробки, проте не здатні виявляти нові або нетипові аномалії.

Підходи, засновані на методах обробки природної мови та глибокому навчанні (LSTM-послідовності, LogBERT, embedding-based кластеризація), розглядають логи як текстові або подієві послідовності та дозволяють виявляти контекстні аномалії [12]. Такі методи є особливо актуальними для задач кореляційного та безпекового аналізу.

Метою даної роботи є підвищення ефективності роботи систем моніторингу за рахунок систематизації існуючих ML-підходів з відкритим кодом, аналіз їх придатності до використання в середовищі Zabbix у трьох взаємозв'язаних напрямках: виявлення аномалій у часових рядах (anomaly detection), прогнозування метрик (forecasting) та аналіз лог-файлів (log analysis), а також формування практичних рекомендацій щодо вибору і застосування відповідних моделей.

Основна частина дослідження

Аналіз наукових досліджень показує про наявність ряду обмежень у сучасних підходах із застосуванням машинного навчання в системах моніторингу. До них відносяться: відсутність уніфікованої послідовності етапів використання ML, орієнтація на внутрішню інтеграцію із Zabbix, що ускладнює створення та масштабування рішень [2, 4]. Перспективним напрямом розвитку є використання сервісу-посередника (ML-gateway), який отримує дані через Zabbix API, виконує ML-аналіз та повертає результати у вигляді елементів даних або тригерів.

Також залишається проблема інтеграції ML-моделей у виробничі системи моніторингу та їх подальший контроль [1]. Одним із можливих шляхів вирішення цих проблем є застосування гібридних підходів, що поєднують використання статистичних моделей з методами глибокого навчання (deep learning) та засобами пояснення результатів (SHAP, feature attribution).

Значна частина досліджень зосереджена на аналізі окремих технічних метрик (CPU, використання пам'яті тощо), а не на оцінці якості сервісу в цілому [9]. Такий підхід ускладнює безпосередній зв'язок результатів аналізу з вимогами до рівня сервісу. Одним із напрямів подальших досліджень є навчання моделей машинного навчання на агрегованих показниках рівня сервісу (SLI) з метою прогнозування ймовірності порушення роботи встановлених SLO.

Формалізований алгоритм вибору ML-моделі

На основі проведеного огляду та порівняльного аналізу ML-підходів процес вибору оптимальної моделі для Zabbix-моніторингу може бути зведений до покрокового алгоритму, орієнтованого на практичне впровадження. На рис.1 наведена загальна схема рішень щодо вибору ML-підходу.

Вхідними параметрами алгоритму є: тип задачі моніторингу; характер і складність даних; наявність сезонності та кореляцій між метриками; критичність сервісу з точки зору SLO; вимоги до затримки обробки та доступні обчислювальні ресурси.

Алгоритм вибору включає послідовне визначення класу задачі, характеристик даних та рівня критичності сервісу з подальшим вибором класу моделей — від статистичних і класичних ML-методів до моделей глибокого навчання. Враховуючи технічні та фінансові обмеження ресурсів використання гібридного підходу, при якому складні для обчислень DL-моделі застосовуються лише для критичних компонентів.

Результатом роботи алгоритму є вибір оптимальної ML-моделі або їх комбінації для конкретної задачі Zabbix-моніторингу.

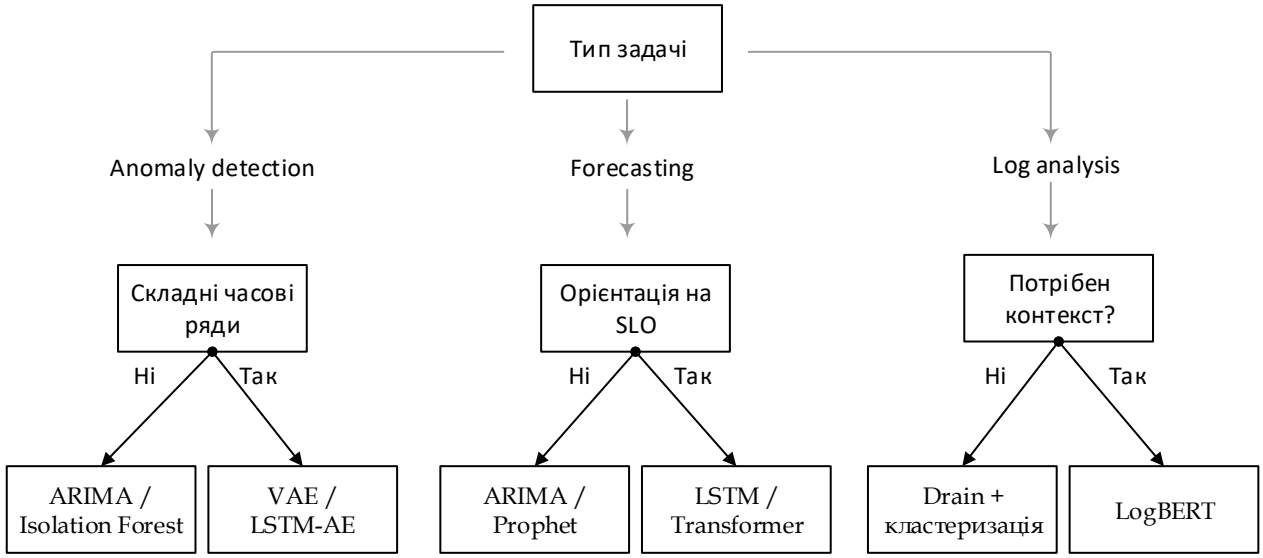

Рис. 1. Decision Tree вибору ML підходу

Узагальнення вибору оптимальних підходів

Незважаючи на значну кількість методів машинного навчання, у даному дослідженні розглядається обмежена множина моделей, що зумовлено специфікою задач IT-моніторингу в середовищі Zabbix та вимогами до практичної інтеграції й експлуатації.

До алгоритму включено лише моделі з відкритим кодом, що мають усталені реалізації у поширених бібліотеках та підтверджену ефективність у сучасних наукових публікаціях. Ключовим критерієм відбору є баланс між точністю, інтерпретованістю та обчислювальною складністю. Інші підходи, зокрема складні ансамблеві, GAN- або foundation-моделі, не розглядалися через надмірні експлуатаційні витрати та обмежену практичну доцільність для задач оперативного IT-моніторингу.

Оптимальний ML-підхід для Zabbix-моніторингу визначається не лише типом задачі, а й критичністю сервісу та експлуатаційними обмеженнями (Табл. 1). Запропонований алгоритм дозволяє систематизувати вибір моделей та забезпечити баланс між точністю, інтерпретованістю і витратами на впровадження.

Таблиця 1

Узагальнена таблиця вибору ML моделі

Тип задачі	*Характер даних*	*Критичність SLO*	*Рекомендована модель*	*Обґрунтування вибору*
Anomaly detection	Прості метрики	Низька	ARIMA / Isolation Forest	Мінімальна затримка, простота реалізації
Anomaly detection	Багатовимірні часові ряди	Висока	VAE / LSTM-AE	Урахування сезонності та кореляцій
Forecasting	Окремі метрики	Середня	ARIMA / Prophet	Інтерпретованість і стабільність
Forecasting	Агреговані SLI	Висока	LSTM / Transformer	Прогнозування ризику порушення SLO
Log analysis	Структуровані логи	Середня	Drain + кластеризація	Масштабованість та швидкодія
Log analysis	Неструктуровані логи	Висока	LogBERT	Контекстний аналіз подій

Методологія ML-орієнтованого моніторингу

1. Загальний підхід та потік даних (Data Flow)

Запропонована методологія ґрунтується на сервісно-орієнтованій інтеграції системи моніторингу Zabbix із зовнішнім аналітичним ML-модулем з відкритим кодом. Zabbix виступає джерелом телеметричних даних та середовищем прийняття рішень, тоді як обчислювально складні алгоритми машинного навчання виконуються у спеціалізованому сервісі аналітики.

Узагальнений потік даних включає такі етапи:

1) збір метрик і логів стандартними механізмами Zabbix;
2) отримання історичних та поточних значень через Zabbix API;
3) попередню обробку даних (очищення, нормалізацію, визначення часових проміжків);
4) передавання даних до ML-модулів (anomaly detection, forecasting, log analysis);
5) обчислення кількісних показників виявлення аномалій, прогнозованих значень або значень, що характеризують ризик порушення SLO;
6) повернення результатів у Zabbix у вигляді нових даних (items) для запуску тригерів і надсилання сповіщень.

Запропонований сервісно-орієнтований підхід відповідає рекомендаціям сучасних досліджень щодо відокремлення ML-логіки від систем збору телеметрії та її інтеграції через API [2, 4].

2. Формалізація алгоритму виявлення аномалій.

Для задач anomaly detection пропонується використання автоенкодерних нейронних мереж, навчених на історичних даних, що відповідають нормальному режиму роботи системи [8]. Нехай $x_t \in R^n$ — вектор значень n метрик у момент часу t.

Автоенкодер реконструює вхідний вектор, формуючи відновлений вектор $\hat{x_t}$. Показник аномальності формально визначається як евклідова норма похибки реконструкції:

$$S_{anom}(t) = || x_t - \hat{x_t} ||_2.$$

Порогове значення θ визначається на основі статистичного аналізу значень S_{anom} у нормальному режимі роботи (за квантильним підходом або правилом μ + kσ). Якщо $S_{anom}(t) > \theta$, то поточний стан системи (елемента, метрики) визначається як аномальний, далі відповідний показник передається до Zabbix для подальшої обробки.

3. Метод прогнозування та оцінки ризику порушення SLO

З огляду на принципи Site Reliability Engineering (SRE), автоматизоване прогнозування ризику порушення SLO є важливим елементом проактивного моніторингу [9, 10]. Для прогнозування динаміки метрик, що формують Service Level Indicators (SLI), застосовуються статистичні та нейронні моделі часових рядів (такі як ARIMA, LSTM або Transformer). Нехай y_t — значення SLI у момент часу t, L — порогове значення, визначене відповідним показником рівня сервісу (SLO).

Прогноз на горизонті h позначається як $\hat{y}_{t+h}$. Ризик порушення SLO формально оцінюється як нормалізоване перевищення прогнозного значення над порогом:

$$R_{slo}(t, h) = \max(0, \hat{y}_{t+h} - L) / L.$$

Отримане значення R_{slo} використовується для формування превентивних тригерів у Zabbix, що дозволяє ініціювати коригувальні дії до фактичного порушення SLO.

4. Методика аналізу логів

Аналіз логів реалізується у два послідовні етапи:

1) Неструктуровані лог-повідомлення перетворюються у впорядковані послідовності подій за допомогою алгоритмів шаблонного парсингу.
2) Сформовані послідовності аналізуються моделлю, яка оцінює ймовірність появи інциденту у заданому контексті.

Показник аномальності в послідовності записів подій формально визначається як від'ємний логарифм умовної ймовірності відстежуваної події:

$$S_{log} = -\log P(e_i \mid e_{i-1}, \ldots, e_{i-k}).$$

Високі значення S_{log} інтерпретуються як можливі інциденти та агрегуються у список подій моніторингу для подальшої обробки в Zabbix.

Запропонована методологія визначає загальні принципи інтеграції ML-компонентів у Zabbix-моніторинг, формалізує потоки даних, алгоритми аналізу та кількісні показники оцінки стану системи. Водночас її практична реалізація потребує чіткого розмежування типів задач моніторингу, відповідних класів ML-моделей і результатів, що повертаються у систему прийняття рішень. З цією метою далі виконано порівняльний аналіз та сформовано класифікаційну схему використання ML-моделей у Zabbix-моніторингу.

Класифікаційна схема використання ML-моделей у Zabbix-моніторингу

Класифікаційна схема використання ML-моделей у Zabbix-моніторингу (Рис. 2) узагальнює запропоновану методологію та відображає логічний поділ задач, потоків даних і відповідних класів моделей. Її метою є формалізація ролі машинного навчання як зовнішнього аналітичного контуру, що доповнює стандартні механізми Zabbix і орієнтується на досягнення SLO.

Класифікаційна схема відображає роль Zabbix як джерела телеметрії та системи прийняття рішень, тоді як ML-моделі формують окремий аналітичний контур для задач anomaly detection, forecasting та log analysis. Запропонована класифікаційна схема забезпечує чітке розмежування функцій між Zabbix і ML-компонентами, спрощує масштабування рішень і створює основу для подальшої автоматизації вибору моделей відповідно до алгоритму вибору ML-моделі.

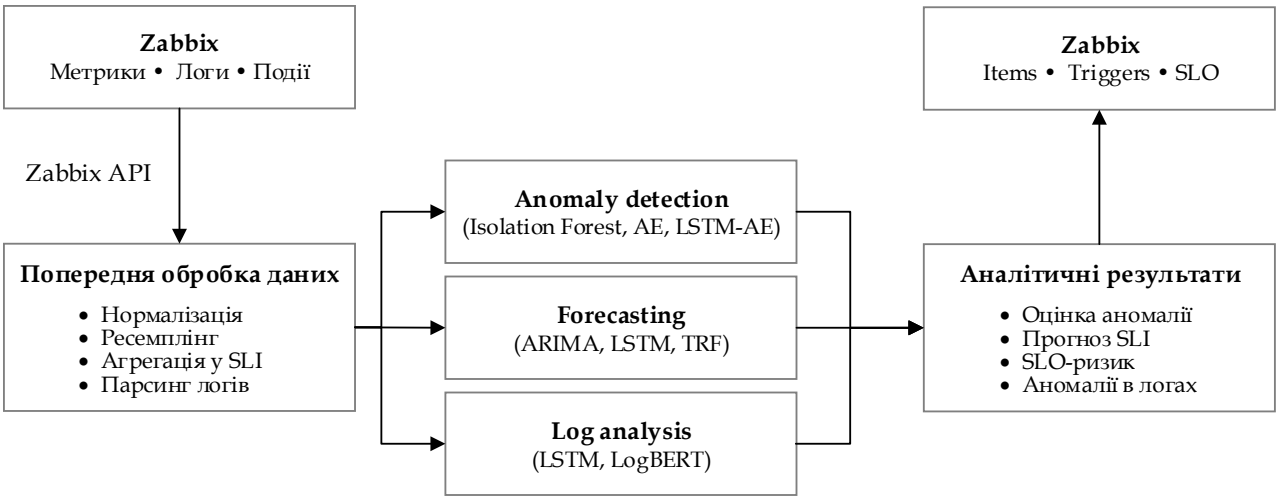

Рис. 2. Класифікаційна схема використання ML-моделей у Zabbix-моніторингу

Порівняльний аналіз ML-методів за типами задач

Порівняльний аналіз ML-методів виконано окремо для кожного типу задач з метою обґрунтованого вибору оптимальних підходів у контексті Zabbix-моніторингу. Аналіз базується на критеріях точності, затримки обробки, інтерпретованості та придатності до інтеграції у виробниче середовище.

Для задач anomaly detection найбільш доцільними є Isolation Forest та Autoencoder-підходи, які забезпечують оптимальний вибір між показниками якості виявлення аномалій і обчислювальною складністю. У задачах forecasting найбільш підходящими є ARIMA та LSTM-подібні моделі в залежності від вимог до точності та об'єму історичних даних. Для log analysis перевага надається комбінації алгоритмів розбору та структуризації текстових записів в логах і нейронних моделей, здатних враховувати контекст подій.

Загальний результат порівняння підтверджує доцільність використання різних ML-підходів в залежності від типу даних і задач, що повністю узгоджується з запропонованим алгоритмом вибору моделей.

Висновки

У роботі проаналізовано сучасні підходи машинного навчання, що застосовуються в системах IT-моніторингу, та обґрунтовано доцільність їх інтеграції в середовище Zabbix для задач виявлення аномалій, прогнозування метрик і аналізу логів. Показано, що стандартні механізми Zabbix не забезпечують повноцінної підтримки сучасних ML-алгоритмів, що зумовлює необхідність використання зовнішніх аналітичних модулів.

Сформовано формалізований алгоритм вибору ML-підходів, який враховує характеристики часових рядів, вимоги до SLO та обчислювальні обмеження. Запропоновано методологію ML-орієнтованого моніторингу, засновану на сервісно-орієнтованій інтеграції через Zabbix API, та класифікаційну схему використання ML-моделей в залежності від типу даних і задач, яка підтверджує необхідність збалансованого використання методів anomaly detection, forecasting і log analysis для підвищення проактивного виявлення інцидентів, а також може бути використана при проєктуванні ML-орієнтованих рішень на основі системи Zabbix.

Література

[1] Zimelewicz E., Biesialska M., Zhang X., Back T., Briand L. Machine Learning-Enabled Systems: Model Deployment and Monitoring Challenges // *Proceedings of the 42nd International Conference on Software Engineering*. – Seoul, Republic of Korea, 2020. – P. 38–49. DOI: 10.1145/3377811.3380394.

[2] Kanavel S., Krishnan S., Wu E., Madden S. Monitoring Machine Learning Systems // *Proceedings of the 2019 IEEE International Conference on Big Data*. – Los Angeles, USA, 2019. – P. 5823–5826. DOI: 10.1109/BigData47090.2019.9005604.

[3] Blázquez-García A., Conde A., Mori U., Lozano J. A. A Review on Outlier/Anomaly Detection in Time Series Data // *ACM Computing Surveys*. – 2023. – Vol. 56, No. 3. – Art. 52. DOI: 10.1145/3573381.

[4] de Souza N., Siegmund N., Apel S., Kästner C. Challenges for Machine Learning Model Integration in Software Systems // *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)*. – 2020. – P. 153–164. DOI: 10.1145/3368089.3409735.

[5] Narayan A., Kumar R., Singh A. AI-Driven Anomaly Detection in Nagios and Zabbix Logs // *International Journal of Computer Applications*. – 2021. – Vol. 174, No. 27. – P. 15–22. URL: https://www.ijcaonline.org/archives/volume174/number27/ (дата звернення: 22.12.2025).

[6] Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice : monograph. – 3rd ed. – Melbourne : OTexts, 2021. – 382 p. URL: https://otexts.com/fpp3/ (дата звернення: 22.12.2025).

[7] Ruff L., Kauffmann J., Vandermeulen R. et al. A Unifying Review of Deep and Classical Anomaly Detection // *Proceedings of the IEEE*. – 2022. – Vol. 110, No. 5. – P. 756–782. DOI: 10.1109/JPROC.2022.3158238.

[8] Blázquez-García A., Conde A., Mori U., Lozano J. A. A Review on Outlier/Anomaly Detection in Time Series Data // *ACM Computing Surveys*. – 2023. – Vol. 56, No. 3. – Art. 52. DOI: 10.1145/3573381

[9] Beyer B., Jones C., Petoff J., Murphy N. R. Site Reliability Engineering: How Google Runs Production Systems. – Sebastopol : O'Reilly Media, 2016. – 552 p. – ISBN 978-1-4919-2922-2.

[10] Lim B., Arik S. O., Loeff N., Pfister T. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting // *International Journal of Forecasting*. – 2021. – Vol. 37, No. 4. – P. 1748–1764. DOI: 10.1016/j.ijforecast.2021.03.012.

[11] He P., Zhu J., Zheng Z., Lyu M. R. Drain: An Online Log Parsing Approach with Fixed Depth Tree // *Proceedings of the 2017 IEEE International Conference on Web Services*. – Honolulu, USA, 2017. – P. 33–40. DOI: 10.1109/ICWS.2017.13.

[12] Guo H., Yuan J., Wu Y., Hu Z. LogBERT: Log Anomaly Detection via BERT // *Proceedings of the 2021 IEEE International Conference on Cloud Computing*. – Chicago, USA, 2021. – P. 1–10. DOI: 10.1109/CLOUD53861.2021.00011.

УДК 004.67

Martyniuk I.V., Okhrimenko T.O. Methodology and classification of open-source ML methods for IT monitoring based on the Zabbix system

Abstract. *In this paper examined the use of open-source machine learning methods for IT monitoring tasks based on the Zabbix system. Analyzed approaches to anomaly detection, time series forecasting, and log file analysis, as well as their limitations in the context of operational monitoring. Proposed a methodology for integrating external ML modules with Zabbix and a classification scheme for using ML models depending on the type of data and needed tasks. Performed a comparative analysis of ML approaches and formulated recommendations for their practical application, taking into account the requirements for achieving the target service level (SLO).*

Keywords: *cybersecurity, information technology, IT-monitoring, Zabbix, machine learning, ML models, anomaly detection, forecasting, log analysis.*

Мартинюк Ігор Вадимович, аспірант Державного університету «Київський авіаційний інститут», інженер ДержНДІ технологій кібербезпеки.
Martyniuk Ihor, PhD student at the State University «Kyiv Aviation Institute», engineer at the State Scientific and Research Institute of Cybersecurity Technologies and Information Protection

Охріменко Тетяна Олександрівна, кандидат технічних наук, старший дослідник, заступник декана з наукової роботи, Державний університет «Київський авіаційний інститут».
Okhrimenko Tetiana, Ph.D., senior researcher, deputy dean for research, State University «Kyiv Aviation Institute».