УДК 004.75 DOI: 10.30837/0135-1710.2025.186.055

В.В. ПРОСОЛОВ, Г.З. ХАЛІМОВ, П.В. ШУЛІК, А.О. СМІРНОВ, Д.О. В'ЮХІН

ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ АТАК НА БЛОКЧЕЙН-СИСТЕМИ

Предметом дослідження є методи виявлення атак у мережах із консенсусом Proof-of-Stake (PoS). Мета роботи – експериментальне дослідження та аналіз ефективності класичних алгоритмів машинного навчання для виявлення шкідливих вузлів у блокчейн-системах. Завдання: аналіз вразливостей технології блокчейн, створення та використання спеціалізованого набору даних для мереж PoS, а також побудова й тестування моделей машинного навчання. Основну увагу приділено порівнянню трьох алгоритмів – Random Forest, Support Vector Machine та k-Nearest Neighbors – для визначення їхньої придатності для моніторингу активності вузлів та виявлення аномалій. Для виконання поставлених завдань застосовано методи: моделювання, емпіричні та математичні методи. Моделювання полягало у програмній реалізації вибраних алгоритмів і подальшому аналізі їхніх результатів із використанням метрик точності, повноти, F1-міри та матриці плутанини. Емпіричні методи реалізовано шляхом тестування моделей на напівсинтетичному датасеті, який містить понад 10 тисяч записів про вузли та транзакції блокчейну. Математичні методи передбачають обчислення статистичних показників ефективності, а також аналіз інформативності ознак, що визначають поведінку вузлів.

Досягнуті результати: апробовано датасет для блокчейнів PoS, що включає ключові операційні параметри транзакцій і вузлів, сформовані пропозиції щодо подальшого використання моделей машинного навчання, здійснено тестування моделей машинного навчання.

Висновки. Доведено що машинне навчання є ефективним інструментом для ідентифікації аномалій та шкідливої активності у блокчейн-системах. Отримані результати закладають підґрунтя для подальших досліджень, які можуть бути спрямовані на розширення ознакового простору, інтеграцію глибоких нейронних мереж, розробку ансамблевих підходів та адаптацію методів до різних типів блокчейнів.

1. Вступ

Технологія блокчейну забезпечує незмінність записів і децентралізовану валідацію транзакцій, давно вийшовши за межі криптовалют до смарт-контрактів і DApps. Попри криптографічну стійкість, публічні мережі залишаються вразливими до низки атак – подвійного витрачання, атак 51% та їх варіантів (Finney, Vector-76), що підтверджується реальними інцидентами (зокрема в Ethereum Classic (ETC), Bitcoin) і призводить до суттєвих збитків [1], [2]. Для підвищення стійкості все активніше застосовують методи машинного навчання (ML) для виявлення аномалій та шкідливої активності в режимі, наближеному до реального часу: від класичних алгоритмів – Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbors (k-NN) – до глибоких моделей (згорткова нейронна мережа (Convolutional Neural Network (CNN), рекурентна нейронна мережа (Recurrent Neural Network (RNN)) та RL-підходів [3], [4], [5]. Окрему увагу приділяють технікам збереження конфіденційності (диференційна приватність (differential privacy, DP), захищені багатосторонні обчислення (Secure Multi-Party Computation, SMPC)), що дозволяють аналіз без розкриття чутливих даних [6].

Глобальна актуальність теми зумовлена стрімкою цифровізацією економіки та зростанням вартості активів, що циркулюють у публічних і консорціумних блокчейн-системах. Масштабування мереж, поява складних DeFi-сценаріїв і смарт-контрактів підвищують площину атак і потенційні системні ризики, тоді як традиційні засоби контролю не забезпечують потрібної швидкодії та пояснюваності. Умови обмеженої

довіри, відсутність централізованого посередника та вимоги до конфіденційності ускладнюють класичний аудит і оперативне реагування. На цьому тлі методи ML дають змогу будувати відтворювані процедури виявлення аномалій і шкідливої активності в режимі, наближеному до реального часу, без втручання у консенсус. Поєднання ML із техніками збереження приватності (DP/SMPC) створює підґрунтя для безпечної аналітики між учасниками мережі. Отже, розроблення надійних, інтерпретованих та масштабованих ML-рішень для моніторингу блокчейн-мереж є важливим глобальним завданням забезпечення кібербезпеки та фінансової стійкості.

2. Огляд сучасних наукових публікацій

Розвиток блокчейн-систем супроводжується зростанням площини атак і появою нових векторів зловмисної активності. Огляди публікацій за останні роки фіксують зміщення акцентів від загальних питань DLT-архітектур до прикладних методів моніторингу та автоматизованого виявлення вторгнень у публічних і консорціумних мережах [7], [8], [9]. Ключова ідея полягає у використанні ML й аналізу аномалій для перетворення реєстрових і телеметричних даних на рішення про ризики в реальному часі, із дотриманням обмежень на приватність і втручання в консенсус.

Класи атак на публічних блокчейнах (double spending, 51 % (Goldfinger), Finney, Vector-76) добре формалізовані в працях із безпеки Bitcoin та суміжних мереж [10], [11], [12]. Практичні інциденти – від компрометації MtGox і DAO до багатоепізодних атак на ETC – засвідчили, що навіть зрілі мережі залишаються чутливими до перехоплення обчислювальної/економічної переваги або до експлуатації специфіки протоколів [1], [2]. На цьому тлі методи раннього виявлення аномалій розглядаються як необхідний додаток до протокольних контрзаходів і економічних стимулів.

Лінія робіт з кібербезпеки на основі ML демонструє ефективність як контрольованих, так і неконтрольованих підходів. Класичні алгоритми – SVM, RF, k-NN – застосовують для виявлення відомих шаблонів зловмисної поведінки та для побудови бінарних класифікаторів транзакцій/вузлів [4], [13]. Стандартні довідники з ML підкреслюють залежність якості таких моделей від інженерії ознак та якості маркованих даних [14]. Для великих вибірок, характерних для блокчейнів, розглядають обчислювально ощадливі варіанти k-NN і прийоми прунінгу, що знижують вартість інференсу без істотної втрати якості [15], [16].

Неконтрольовані методи та детектори аномалій актуальні там, де марковані дані обмежені або атаки – «нульового дня». Узагальнюючий огляд [9] систематизує застосування кластеризації (k-means, DBSCAN) та моделей на кшталт Isolation Forest/one-class підходів для виявлення рідкісних відхилень у поведінці вузлів і потоках транзакцій. Перевага цих підходів – пластичність до нових патернів; обмеження – потреба в тонкому тюнінгу порогів і чутливість до зміни середовища.

Глибоке навчання все частіше інтегрують у стек безпеки блокчейнів. Роботи з кібербезпеки демонструють переваги CNN/RNN для аналізу структурованих графів транзакцій і часових рядів активності [3]. У прикладних сценаріях (енергетичні мережі, обмін активами) запропоновані фреймворки на стику машинного навчання і блокчейну показують підвищення чутливості до аномалій за прийнятної затримки [5]. Для індустріального IoT перспективними визнані AI-механізми довіри та адаптивні політики реагування, що поєднують глибокі моделі з блокчейн-верифікацією [17].

Окрему нішу займають генеративні моделі. Generative Adversarial Networks (GAN) використовують для отримання реалістичних синтетичних сценаріїв атак, що розширюють навчальні вибірки детекторів і дозволяють відпрацьовувати рідкі події без появи небезпечних кейсів у продакшені [18]. Недоліки – висока обчислювальна вартість і

залежність від якості базових даних.

Питання конфіденційності та регуляторної сумісності стимулюють застосування методів збереження приватності (DP, SMPC). Децентралізовані протоколи аналізу, які не потребують розкриття сирих даних, доводять практичність інтеграції аналітики в чутливі домени – наприклад, охорона здоров'я й фінанси на блокчейні [6]. Такі методи є ключовими для PoS/DPoS-мереж, де відкритий агрегаційний моніторинг може конфліктувати з вимогами приватності валідаторів.

На рівні протоколів триває переосмислення властивостей консенсусу. Від класичної моделі BFT до практик PoS/DPoS –у наукових публікаціях підкреслено тонкі компроміси між живучістю, безпекою та продуктивністю [19], а також пропонується покращення голосувальних механізмів і стейкінгових правил для зниження ймовірності маніпуляцій [20]. Для PoS-ланцюгів окремо виділяють довготривалі (long-range) атаки й методи їхнього виявлення за допомогою глибоких моделей, навчених на історії реєстру та поведінці валідаторів [21].

Нарешті, якість емпіричних досліджень суттєво залежить від наявності репрезентативних датасетів. Через складність повного доступу до ончейн/офчейн телеметрії поширеною практикою стає створення напівсинтетичних наборів, що поєднують витягнуті з експлорерів ознаки з реалістичними симуляціями для масштабування корпусу спостережень [22]. Такі датасети дозволяють відтворювано порівнювати алгоритми (RF, SVM, k-NN) й валідувати гіпотези щодо інформативності конкретних показників (комісії, вікові метрики монет, щільність блоків, оцінки/ваги тощо).

Підсумовуючи, сучасні наукові публікації демонструють рух від «модель-центричних» до «системоцентричних» підходів. Це поєднання класичних ML-алгоритмів, глибоких моделей, генеративних симуляторів і технік приватності в єдиний конвеєр моніторингу. Водночас зберігаються прогалини у відкритих PoS-даних, у стандартизації ознак і в пояснюваності глибоких детекторів. Саме тут доречні деревні ансамблі з оцінкою важливостей ознак, поєднані з обережно сконструйованими напівсинтетичними корпусами – як базис для практичних систем раннього виявлення загроз.

У публічних PoS-блокчейнах бракує відтворюваного та операційно придатного ML-конвеєра для виявлення шкідливих вузлів, який працює без втручання у протокол, спирається лише на ончейн-ознаки, надає калібровані ймовірності з пороговим вибором з урахуванням вартості помилок та забезпечує пояснюваність (інформативність ознак) і статистично коректну валідацію. Існуючі роботи часто не узгоджують препроцесинг і валідацію, рідко враховують коштовність FN/FP, а також не дають практичних правил вибору робочої точки для моніторингу мережі.

3. Мета і задачі дослідження

Мета дослідження – експериментальне дослідження та аналіз ефективності класичних алгоритмів машинного навчання для виявлення шкідливих вузлів у блокчейн-системах.

Таким чином, в ході дослідження необхідно вирішити такі задачі:

– сформувати відтворювану постановку задачі виявлення шкідливих вузлів у PoS-блокчейні на основі доступних ончейн-ознак, із чітким поділом даних на тренувальну та тестову частини;

– реалізувати уніфікований пайплайн «препроцесинг → модель» для класичних алгоритмів (RF, SVM з RBF-ядром, k-NN) з урахуванням вимог до масштабування ознак і відтворюваності;

– виконати порівняльне моделювання обраних алгоритмів із підбором гіперпараметрів і валідацією на відкладеній вибірці;

– оцінити якість класифікації за узгодженим набором метрик (Accuracy, Precision, Recall, F1, PR-AUC) та проаналізувати типові помилки за матрицями плутанини;

– забезпечити пояснюваність результатів через аналіз інформативності ознак та виділення ключових індикаторів ризику для подальшого моніторингу мережі;

Очікуваним результатом дослідження є порівняльна оцінка ефективності класичних алгоритмів (RF, SVM з RBF-ядром, k-NN) з вибором робочих режимів (порогів) для різних профілів ризику, узгоджених із компромісом між чутливістю та специфічністю; формування переліку найінформативніших ознак; а також підготовка рекомендацій щодо інтеграції перевірених конфігурацій моделей і порогових налаштувань у процедури безперервного моніторингу блокчейн-мереж.

4. Матеріали і технології дослідження

4.1. Об'єкт та основна гіпотеза дослідження

Об'єктом дослідження є вузли публічної PoS-блокчейн-мережі та їхні поведінкові й операційні ознаки, за якими ідентифікується шкідлива активність.

Дослідження спрямоване на побудову цілісного технологічного ланцюга – від формування придатної для аналізу вибірки та інженерії ознак до навчання моделей і верифікації їхньої здатності відрізняти шкідливу активність від доброчесної. Тому основною гіпотезою даного дослідження є гіпотеза, за якою сукупність операційних і поведінкових індикаторів вузлів та транзакцій (комісії, параметри стейкінгу, «вік монети», щільність і оцінка блоку, структура розміру транзакцій тощо) містить достатньо інформації для надійної бінарної класифікації вузлів без втручання в протокольні процеси та без порушення конфіденційності. Для перевірки цієї гіпотези було запропоновано провести експериментальні дослідження з використанням напівсинтетичного датасету PoS на 10 000 записів з чітко визначеною цільовою змінною. В межах цих досліджень було запропоновано виконати повну підготовку даних (очищення, нормування для чутливих до масштабу алгоритмів) і формування відтворюваної процедури поділу вибірки на тренувальну та тестову (train/test).

4.2. Технологія блокчейну та аналіз її вразливості

Блокчейн розглядається як послідовність блоків $\mathcal{B} = \{B_0, B_1, \dots, B_t\}$, де B_0 – генезис-блок. Кожен блок $B_i = (H_i, T_i)$ складається із заголовка H_i та множини транзакцій T_i. У заголовку зберігаються ключові поля: version, prev_block $= h(H_{i-1} \parallel T_{i-1})$, merkle_root $= \mu(T_i)$, time, bits, nonce. Незмінність досягається через хеш-вказівник prev_block: будь-яка модифікація попередніх даних змінює хеші всіх наступних блоків. Для підтвердження включення транзакції tx_j використовується мерклеве дерево: лист $l_j = h(tx_j)$, внутрішні вузли $n_j^{(k-1)} = h(n_{2j-1}^{(k)} \parallel n_{2j}^{(k)})$, корінь $\mu(T_i)$ фіксується в заголовку. Доказ включення має довжину $O(\log m)$ і не потребує сканування всього блоку [7].

Моделі стану бувають двох типів. У Bitcoin транзакція «спалює» набір незатрачених виходів і створює нові; валідність перевіряється підписами та відсутністю дублювання витрати. В account-based (ETC) підтримується глобальний стан $\mathcal{S}$ рахунків/контрактів; транзакція змінює баланси, код або сховище. Для верифікації стану застосовується Merkle–Patricia Trie: у заголовку зберігається корінь стану (state root), який дає можливість формувати короткі пруфи наявності/значення рахунку. Така організація забезпечує ідентифікацію мінімальних змін і підвищує ефективність верифікації.

Механізми консенсусу визначають, як блоки додаються до ланцюга і яка гілка вважається канонічною. У Proof-of-Work (PoW) майнер шукає nonce, що задовольняє умові

$$h(H_i) < \text{target} = 2^{\lambda}/D, \tag{1}$$

де D – складність; λ – довжина хешу.

При цьому діє правило найдовшого/«найважчого» ланцюга; імовірність реорганізації зменшується зі зростанням кількості підтверджень k_{conf} [23].

У Proof-of-Stake (PoS) валідатор для слоту/епохи обирається пропорційно стейку s_v:

$$\mathbb{P}\{\text{validator} = v \mid \mathcal{E}\} = s_v / \sum_u s_u \,. \quad (2)$$

Для PoS характерні механізми фіналізації (комітети/порогові голосування), що ускладнюють глибокі реорганізації та вводять економічні покарання (slashing) за зловмисні дії.

Схему блокчейну наведено на рис. 1.

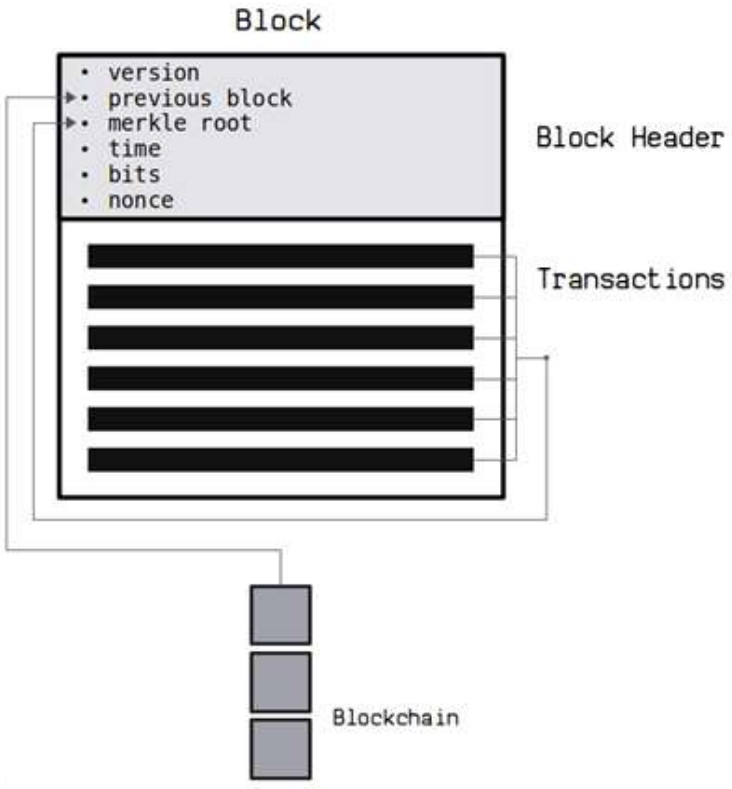

Рис. 1 Схема блокчейну

На рівні протоколу та мережі існує низка типових загроз. Подвійне витрачання (Double Spending) у PoW полягає в просуванні приватної гілки, яка містить альтернативну транзакцію. Якщо зловмисник має частку ресурсу $q < p$ (де $p = 1 - q$), то ймовірність «наздогнати» легальний ланцюг експоненційно спадає з k_{conf} - кількістю підтверджень; практичний контрзахід – чекати достатньо великий k_{conf}, пропорційний вартості платежу [23]. Finney-атака [10] передбачає попередній видобуток блоку з депозитом на свою адресу та його публікацію після отримання товару за альтернативною транзакцією; ефективний контрзахід – чекати кілька підтверджень. Vector 76 [11] – комбінована атака на біржі, коли приватна гілка з депозитом розсилається напряму вузлам біржі одночасно із запитом на вивід; за певних умов біржа приймає гілку з депозитом як канонічну, а далі втрачає кошти після реорга. Атака 51 % (Goldfinger) [11] можлива за контролю > 50 % ресурсу (хеш-потужності або ефективного стейку), що дає змогу нав'язати історію, цензурувати транзакції та систематично проводити подвійні витрати; у PoS економіка штрафів і фіналізація підвищують ціну атаки. Окремо виділяються вразливості смарт-контрактів (реентрансі, порушення інваріантів, некоректні перевірки), що здатні призводити до втрати активів без порушення базового консенсусу [19] – тут ефективні формальна верифікація та незалежні аудити.

Практична політика підтверджень/фіналізації має враховувати вартість транзакції, ризик середовища та консенсус. Для PoW дрібні платежі можуть обходитись 1-2

підтвердженнями, великі – вимагати 6+; для PoS доцільно очікувати фіналізації чекпойнтів. Біржам та сервісам рекомендовано [23] застосовувати затримки на депозити/виводи (проти Vector 76), використовувати георознесені вузли з незалежними каналами підключення, моніторити показники форків (orphan rate, head-дисонанси), а також сповіщення про підозрілі патерни пропагації блоків. Для PoS економічні стимули/штрафи (slashing) разом із прозорими правилами вибору гілки (GHOST/LMD-GHOST) та фіналізацією (на кшталт Casper FFG) зменшують простір атак і підвищують вартість їх реалізації.

4.3. Застосування методів та алгоритмів машинного навчання для забезпечення безпеки блокчейну

Методи аналізу даних посідають вагоме місце в кібербезпеці [12], а поширення сучасних підходів ML суттєво підвищило точність і оперативність виявлення загроз як у реальному часі, так і на етапі пост-інцидентного аналізу [13]. Для блокчейн-систем, де події формують безперервний потік реєстрових записів і телеметрій, практично поєднують дві комплементарні парадигми: контрольоване навчання для розпізнавання відомих патернів атак і неконтрольовані методи для пошуку відхилень, включно з «нульовим днем» [9]. У задачах класифікації транзакцій або вузлів на «легітимні» та «шкідливі» ключову роль відіграють SVM і RF [1], [4]. SVM завдяки використанню ядер коректно працює з нелінійними межами рішень, але відчутна складність навчання й інференсу на дуже великих вибірках поєднується з нижчою інтерпретованістю. RF як ансамбль дерев рішень краще масштабується на високу розмірність, є стійким до шуму й надає важливості ознак, що зручно для операційного моніторингу, хоча прозорість моделі нижча, ніж у поодинокого дерева. Часто використовують і k-NN як просту відправну точку: він чутливий до локальних відхилень розподілу, але потребує уважного масштабування ознак і має дорогий інференс на великих даних. Загалом продуктивність контрольованих підходів критично залежить від якості маркованих наборів і доменної інженерії ознак; репрезентативні приклади класів «атака/норма» та релевантні характеристики (комісії, параметри стейку, «вік монети», щільність/оцінка блоку тощо) визначають фактичну межу досяжної якості [20].

Коли еталонних міток бракує або в середовищі швидко з’являються нові вектори атак, на перший план виходять неконтрольовані підходи [9], [14]. Кластеризація на кшталт k-means або DBSCAN групує подібні транзакції/вузли і допомагає виявляти нетипову поведінку без попереднього знання її природи: k-means приваблює швидкодією на великих масивах, але припускає «сферичність» кластерів і чутливий до ініціалізації, тоді як DBSCAN не потребує фіксувати кількість кластерів і здатен виокремлювати викиди, натомість вимагає ретельного добору параметрів і гірше масштабується у дуже високій розмірності. Для безпосереднього пошуку аномалій застосовують Isolation Forest та One-Class SVM: перший ізолює викиди короткими шляхами в випадкових деревах, другий моделює «нормальний» клас і позначає відхилення як підозрілі [4]. Тут особливо важливі коректне масштабування, вибір метрик відстані й калібрування порогів спрацювання, оскільки інакше зростає частка хибних спрацювань.

За наявності великих і різноманітних даних доцільно залучати методи глибокого навчання [5]. CNN зручно працюють з тензорними поданнями структур (наприклад, перетвореннями графів транзакцій на матриці/«зображення»), тоді як RNN моделюють часові залежності в послідовностях транзакцій і виконання смарт-контрактів. Такі моделі часто досягають високої точності, автоматично видобуваючи ознаки, проте «платою» стають значні обчислювальні ресурси й нижча інтерпретованість «чорних скриньок», а також вимоги до захисту приватності під час збирання/навчання. Паралельно досліджують навчання з підкріпленням: агенти на основі Q-learning та споріднених методів здатні

виробляти адаптивні політики захисту, підлаштовуючись до динамічних загроз у симульованому середовищі мережі, але навчання є складним, ресурсоємним і потребує безпечних полігонів [17].

Окремий клас становлять байєсівські моделі та методи генерації даних. Байєсівські мережі описують причинно-ймовірнісні залежності, витримують неповноту даних і надають прозорі висновки щодо ризиків, що підсилює пояснюваність для інцидент-респонсу й аудиту [14], [24]. GAN застосовують для синтезу реалістичних сценаріїв атак і збагачення тренувальних наборів детекторів, але такі моделі вимогливі до обчислювальних ресурсів, якості вихідних даних і стабільності навчання [18]. Питання приватності вирішують поєднанням диференціальної приватності та захищених багатосторонніх обчислень: DP вводить контрольований шум для гарантій приватності зі зниженням точності, тоді як SMPC дозволяє спільне навчання без обміну сирими даними, проте має суттєві накладні витрати [6].

Підсумовуючи, практично виправдано комбінувати інтерпретовані контрольовані моделі (RF, SVM) на маркованих даних із неконтрольованими детекторами (Isolation Forest, One-Class SVM, DBSCAN) для виявлення раніше невідомих патернів, а за наявності великих графових і послідовнісних корпусів – доповнювати їх CNN/RNN. Вибір конкретної техніки визначають доступність еталонних міток, вимоги до пояснюваності, обчислювальні обмеження та політики приватності [20]. Узагальнене порівняння характеристик і компромісів ML-підходів наведено в табл. 1, яку доцільно використовувати як орієнтир для вибору методу під конкретні умови моніторингу.

4.4. Набір даних блокчейну Proof-of-Stake

Наступний етап дослідження передбачає використання набору даних [22], спеціально розробленого для блокчейнів PoS. PoS – метод захисту у криптовалютах, при якому ймовірність формування учасником чергового блоку в блокчейні пропорційна частці криптовалюти, яку має даний учасник від її загальної кількості. Даний метод є альтернативою методу PoW, при якому ймовірність створення чергового блоку вище у володаря потужнішого обладнання. Через відсутність справжнього набору даних, що охоплює всі етапи створення та перевірки блоків у блокчейнах PoS, було визначено необхідність синтезу такого набору даних. Набір даних було створено шляхом вилучення відповідних ознак, пов’язаних із конкретними етапами, до яких є доступ [21]. Цей набір даних містить інформацію про більш ніж 10000 записів даних вузлів блокчейну PoS. При цьому спочатку було розроблено напівсинтетичний набір даних для додатків на основі блокчейну PoS з 303 записами на основі виходів вузла (валідатора) та блоку (транзакції). Потім цей набір даних було розширено за допомогою бібліотек Python для імітації додаткових записів і для створення розширеного набору даних вже з 10000 записів. Цей набір даних розподілено також між шкідливими та нешкідливими вузлами. Ця методологія для створення набору даних розроблена з метою надати вичерпний і репрезентативний набір даних для використання при аналізу та оцінки систем блокчейну PoS. Останні десять записів цього набору показано на рис. 2.

Якщо подивитися на останній стовпець, можна побачити, що набір даних розподілено між шкідливими вузлами (1) та нешкідливими вузлами (0). Існує деяка невідповідність у співвідношенні звичайних та підозрілих вузлів. Не завжди підозрілі вузли є по-справжньому небезпечними з погляду атак. Таким чином, слід врахувати якусь ймовірнісну невизначеність при виведенні рекомендацій.

Таблиця 1

Порівняння характеристик і компромісів ML-підходів

Алгоритм	Основна мета / задачі	Переваги	Обмеження	Виявлення невідомих атак	Обчислювальні витрати
SVM	Бінарна класифікація транзакцій/вузлів (законна vs шкідлива), IDS/IPS.	Оптимальна межа рішень; добре працює з нелінійностями (ядра).	Масштабування погіршується на дуже великих даних.	Обмежене (краще для відомих патернів на мічених даних).	Середні /високі (залежно від ядра та розміру даних)
RF	Класифікація / регресія; виявлення фроду; виявлення шкідливих вузлів.	Стійкий до шуму; працює з високою розмірністю; оцінка важливості ознак.	Менш прозорий за одне дерево.	Обмежене (в основному вчиться на історичних ярликах).	Середні
k-NN	Виявлення аномалій за відстанями; базова класифікація.	Простота; локальна чутливість; добре ловить відхилення в розподілі.	Повільний на великих наборах (інференс); чутливий до вибору метрики та масштабу; «прокляття розмірності».	Добре, якщо ознаки відображають відстані від норми.	Високі на етапі передбачення
k-means	Групування подібних транзакцій/вузлів; сегментація поведінки.	Швидкий та простий; добре працює на великих наборах даних.	Потребує вибору k; припускає «сферичні» кластери; чутливий до ініціалізації; слабко обробляє шум.	Виявляє невідомі групи/патерни (не мітить аномалії явно).	Низькі / середні
DBSCAN	Кластери довільної форми та виділення викидів.	Не вимагає k; знаходить кластери різної форми; явно позначає викиди.	Складний підбір параметрів на неоднорідних даних; гірше масштабується на дуже високій розмірності.	Добре для пошуку аномалій (потенційно «нульовий день»).	Середні / високі (залежить від індексації відстаней)
Isolation Forest	Виявлення статистичних викидів/аномалій.	Швидкий; працює на високій розмірності; відносно інтерпретований (шляхи ізоляції).	Параметри впливають на чутливість; статична модель (потрібен ретренінг).	Висока придатність до невідомих відхилень.	Низькі/середні
One-Class SVM	Моделювання «нормального» класу та виявлення відхилень.	Потужний для нелінійних розподілів; не потребує прикладів атак.	Погано масштабується.	Високе за умови якісної моделі «норми».	Високі на великих наборах

Кінець таблиці 1

Алгоритм	Основна мета / задачі	Переваги	Обмеження	Виявлення невідомих атак	Обчислювальні витрати
CNN	Аналіз структурованих представлень (наприклад, графи транзакцій як тензори/матриці).	Автоматичне видобування ознак; висока точність; робота з великими даними.	Мала інтерпретованість («чорна скринька»); висока вартість тренування; конфіденційності даних.	Потенційне, за рахунок узагальнення, але залежить від різноманіття даних.	Високі
RNN	Аналіз послідовностей (історія транзакцій, виконання контрактів).	Моделює часову залежність; виявляє послідовні патерни атак.	Складність навчання/довгих залежностей; «чорна скринька»; висока вартість.	Може виявляти нові послідовні аномалії після адаптації.	Високі
Q-learning, інші RL-агенти	Адаптивні політики безпеки; прийняття рішень у реальному часі.	Адаптивність; здатність реагувати на нові загрози; оптимізація дій вузлів.	Складне та ресурсомістке навчання; ризик небезпечних дій під час навчання; потреба у безпечних симуляціях.	Високе (вчиться реагувати на нові патерни).	Високі
Байєсівські мережі	Оцінка ризиків і залежностей; робота з невизначеністю/неповними даними.	Прозорі причинно-наслідкові зв'язки; обробка пропусків; гнучкі до змін.	Складність побудови/масштабування на великому числі змінних; чутливість до припущень.	Може інферувати підвищені ризики без прямих прикладів атак.	Середні
GAN	Генерація реалістичних синтетичних даних атак; підсилення	Покриття сценаріїв атак; тестування/стрес-тести; збагачення тренувальних даних.	Складне навчання (mode collapse тощо); великі витрати ресурсів; залежність від якості даних.	Непряме: допомагає моделювати/симулювати нові атаки.	Високі
DP, SMPC	Захист конфіденційності під час аналізу/навчання; відповідність регуляціям.	Зберігають приватність та анонімність; дозволяють колективний аналіз без витоків.	DP знижує точність через шум; SMPC має великі накладні витрати і складність впровадження.	Опосередковане: дають змогу безпечного обміну даними для кращих моделей.	Високі (особливо SMPC)
Decision Tree	Базова класифікація/правила; пояснювані пороги/умови.	Дуже інтерпретований; швидкий; просте розгортання.	Схильний до перенавчання; точність нижча за ансамблеві методи.	Обмежене.	Низькі

	BlockHeight	UnixTimestamp	TxnFee(ETH)	TxnFee (Binary)	Status (Tags)	Block Generation Rate	Stake Reward	Coin Stake	Stake Distribution Rate	Txnsize	Coin Days	Coin Age	Block Density (%)	Block Score	Coin Day Weight	Node Label
9990	5688453	1529044536	0.000000	1	0	0	0	35	1024	53	3	61	2574	2016	23	0
9991	14846283	1658941589	0.000000	0	0	0	1	69	1424	61	1	118	1441	5353	54	0
9992	5905747	1524145611	0.000000	1	0	0	1	63	1458	34	2	84	1674	1263	0	0
9993	5468684	1650512986	0.516596	1	1	1	1	112	2643	31	1	172	1620	4605	26	0
9994	15450240	1652741638	0.462925	1	1	1	1	189	2734	46	3	164	1577	4211	54	1
9995	6488560	1524145611	0.000000	1	0	0	1	56	934	73	1	41	1455	1430	24	0
9996	5990292	1524145611	0.000000	1	0	0	1	47	1056	26	3	88	1405	1507	30	0
9997	15274922	1660615336	0.482471	1	1	1	1	94	1917	76	1	104	2347	5859	58	1
9998	15450240	1524145611	0.000000	1	1	0	1	32	955	69	1	94	1378	2503	0	1
9999	6085840	1525115380	0.000000	0	0	0	1	55	1746	85	1	81	1217	2915	26	0

Рис. 2. Фрагмент набору даних для блокчейну Proof-of-Stake

Було зібрано такі характеристики цього набору даних: TxHash, висота блоку, UnixTimestamp, TxFee (ETH), TxFee (Binary), Status (Tag). Інші характеристики, виявлені в ході минулих досліджень, включають API-інтерфейси оглядача блоків та синтез минулих досліджень, а саме: швидкість генерації блоків, ставка монети, швидкість розподілу ставок, винагорода за ставки, вік монети, дні монети, щільність блоків, оцінка блоку, вага монети в день, розмір транзакції та node label. Нижче наведено розшифрування деяких параметрів із рис. 2.

Block-generation rate (швидкість генерації блоків) – швидкість, з якою блоки створюються у кожну епоху, починаючи з генезісного блоку.

TxHash – хеш транзакції, пов'язаний із перевіреним блоком.

Coin stake (ставка монет) – кількість монет або токенів, які вузол повинен поставити, щоб бути обраним як валідатор. Мінімальна ставка для нашого дослідження складала 30 токенів. Будь-який вузол, який ставить менше 30 токенів, не буде обраний як валідатор, а ті, хто ставить більше 30, мають вищу ймовірність бути обраними як валідатор.

Stake distribution rate (швидкість розподілу ставок) —процес активної участі у перевірці транзакцій у мережі PoS.

Stake reward (нагорода за ставку) —нагорода в монетах, яку валідатор отримує за перевірку транзакції та додавання її до блоку для створення дійсного блоку. Нагорода буде однаковою для кожного створеного блоку. У створеному наборі Stake reward може приймати значення 1 (Нагорода за ставку) та 0 (Нагорода без ставки).

Transaction fees (комісії за транзакції) різняться залежно від розміру та швидкості транзакції. Для цього набору даних характеристика Transaction fees може приймати значення 1 (комісія є) та 0 (комісії немає).

Coin days (дні монети) —буде представлено у періодах на основі протоколів та правил PoS. 0 днів = 0, (0 – 30] днів = 1, (30 – 60] днів = 2, (60 – 90] днів = 3.

Block height (висота блоку) розраховується шляхом додавання Епохи + Швидкість генерації блоків + Розподіл частки + Нагорода за ставку.

Transaction size (розмір транзакції) – кількість транзакцій у блоці.

Node Label (мітка вузла) позначає шкідливі та нешкідливі вузли.

Як видно, цей набір даних був підготовлений для проведення експериментів з використанням методів ML, оскільки виділена бінарна цільова змінна. Отже, ми можемо на основі цього набору провести бінарну класифікацію або навчити нейронну мережу.

У датасеті мітка вузла (Node Label) визначає, чи є вузол шкідливим (1) або нешкідливим (0). Однак причинно-наслідкові зв’язки для того, щоб вузол був позначений як шкідливий або нешкідливий, базуються на аналізі такого набору характеристик:

Transaction Fee (TxnFee). Високі або аномальні комісії за транзакції можуть бути ознакою підозрілої активності.

Stake Distribution Rate і Stake Reward. Непропорційно великі ставки або винагороди можуть вказувати на шахрайські дії.

Block Density і Coin Age. Непослідовності у щільності блоків або віці монет можуть свідчити про спроби маніпулювати системою.

Block Score. Високий або низький бал блоку, який відхиляється від норми, може бути індикатором потенційних атак.

Інші характеристики. Патерни в поведінці вузлів, наприклад, занадто часті або специфічні транзакції, можуть сигналізувати про шкідливу активність.

5. Опис ходу та результатів дослідження

Вирішувалася задача бінарної класифікації шкідливих вузлів у PoS-блокчейні. Для кожного об'єкта було задано вектор ознак $\mathrm{x} \in \mathbb{R}^d$ і мітку $y \in \{0,1\}$ (де $y = 1$ означає «шкідливий»). Необхідно було побудувати оцінювач ймовірності та порогове рішення:

$$\hat{p}_\theta(\mathbf{x}) \approx \mathbb{P}(Y = 1 \mid \mathbf{x}), \tag{3}$$

$$\hat{y}_t = \mathbf{1}\{\hat{p}_\theta(\mathbf{x}) \geq t\}, t \in [0,1]. \tag{4}$$

Для дослідження використано напівсинтетичний набір із 10000 спостережень. В цьому наборі наведено значення16 числових ознак (в тому числі TxnFee(ETH), Block Generation Rate, Stake Reward, Block Score, Coin Age) та бінарної цільової змінної Node Label. Поділ на train/test склав 70/30 (stratify, random_state = 42). Частки класів склали: приблизно 42 % ($y = 1$) та 58 % ($y = 0$). Пропусків не виявлено.

Масштабування застосовувалося лише для методів, чутливих до масштабу (SVM, k-NN), і виконувалося всередині CV-pipeline на train:

$$z = \frac{x-\mu}{\sigma}, \tag{5}$$

де x – значення ознаки; μ – середнє ознаки; σ – стандартне відхилення. Значення цих параметрів було обчислено на train.

Для деревних методів (RF) масштабування непотрібне. Така схема унеможливлює витоки інформації (data leakage).

Порівнювалися алгоритми RF, SVM з RBF-ядром (SVM (RBF)) та k-NN. Для кожної моделі побудовано sklearn-pipeline «препроцесинг → модель». Добір гіперпараметрів виконувався K-fold стратифікованою крос-валідацією на train (K = 5).

Першим досліджувався алгоритм SVM (RBF).

Його ядро:

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \parallel \mathbf{x} - \mathbf{x}' \parallel_2^2). \tag{6}$$

Оптимізаційна задача (м'які відступи) для даного алгоритму:

$$\min_{\mathbf{w}, b, \boldsymbol{\xi}} \frac{1}{2} \parallel \mathbf{w} \parallel_2^2 + C \sum_{i=1}^n \xi_i \quad \text{s.t. } y_i(\mathbf{w}^\top \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \xi_i \geq 0. \tag{7}$$

Сітка гіперпараметрів мала вигляд:

$$C \in \{0.1, 1, 10\}, \qquad \gamma \in \left\{\frac{1}{d}, \frac{1}{2d}, 0.01\right\},$$

де d – кількість ознак.

Другим досліджувався алгоритм RF.

Використовувався ансамбль із B дерев з випадковими підпросторами ознак; критерій розщеплення – зменшення нечистоти Джині:

$$\Delta G = G(S) - \sum_{j\in\{\text{left,right}\}} \frac{|S_j|}{|S|} G(S_j),\ G(S) = 1 - \sum_{c\in\{0,1\}} p_c^2 . \quad (8)$$

Сітка гіперпараметрів мала вигляд:

$$n_{\text{estimators}} \in \{100, 300, 500\},$$

$$\max\_depth \in \{\text{None}, 10, 20\},$$

$class\_weight =$ 'balanced' .

Масштабування для RF не потрібне.

Третім розглядався алгоритм k-NN.

Для нього правило більшості серед kнайближчих сусідів мало вигляд:

$$\hat{y}(\mathbf{x}) = \arg\max_{c\in\{0,1\}} \sum_{i\in\mathcal{N}_k(\mathbf{x})} \mathbf{1}\{y_i = c\}. \quad (9)$$

Сітка гіперпараметрів мала вигляд:

$$k \in \{5, 11, 15, 21\},\ \text{metric} \in \{\text{euclidean, cosine}\}.$$

Порогове рішення (2) задає компроміс між Recall і Precision: зменшення t підвищує Recall (менше FN) ціною зростання FP; збільшення t підвищує Precision.

Для асиметричних витрат помилок оптимальний поріг позначимо як t^* і визначається:

$$t^* = \frac{c_{FP}}{c_{FP}+c_{FN}}, \quad (10)$$

де c_{FP}— «ціна» хибнопозитивної помилки; c_{FN}— «ціна» хибнонегативної помилки.

PR-крива будувалася як траєкторія $(\text{Recall}(t), \text{Precision}(t))$ при $t \in [0,1]$. Площа під кривою (PR-AUC) може оцінюватися методом трапецій:

$$\text{PR} - \text{AUC} \approx \sum_{j=1}^{m-1} \left(\frac{P_j+P_{j+1}}{2}\right)(R_{j+1} - R_j), \quad (11)$$

де $R_j = \text{Recall}(t_j)$; $P_j = \text{Precision}(t_j)$; $t_1 < \cdots < t_m$.

В процесі порівняння використовувалися такі метрики:

$$\text{Precision} = \frac{TP}{TP+FP}, \quad (12)$$

$$\text{Recall} = \frac{TP}{TP+FN}, \quad (13)$$

$$F_1 = 2 \cdot \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}, \quad (14)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} . \quad (15)$$

де TP – істинно позитивні результати: кількість прикладів, що насправді $y = 1$ і модель передбачила 1, FP – хибно позитивні результати: кількість прикладів, що насправді $y = 0$, але модель передбачила 1, FN – хибно негативні результати: кількість прикладів, що насправді $y = 1$, але модель передбачила 0, TN - істинно негативні результати: кількість прикладів, що насправді $y = 0$ і модель передбачила 0.

Для парного порівняння двох моделей на тих самих прикладах застосовувався тест Мак-Немара. Нехай n_{01} – кількість випадків, де модель А помилилась, а модель В – ні; n_{10} – навпаки. Статистика з поправкою Єйтса:

$$\chi^2 = \frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}} \quad (16)$$

(без поправки: $\chi^2 = \frac{(n_{01}-n_{10})^2}{n_{01}+n_{10}}$). Значущість оцінювалася за χ^2з 1 ступенем вільності (p-value).

Невизначеність точкових оцінок для F1/Recall оцінювалася бутстрепом (семплювання з поверненням, $n = 1000$, рівень довіри 95 %).

Основні особливості порівнюваних алгоритмів наведено в табл. 2.

Таблиця 2

Основні особливості порівнюваних алгоритмів

Алгоритм	Нормалізація даних	Крос-валідація	Матриця плутанини	Контроль витоків (pipeline)	Особливі параметри
RF	Ні	Так (5 фолдів)	Так	Так (CV-pipeline; test лише для фіналу)	Кількість дерев (100)
SVM (RBF)	Так (StandardScaler у pipeline)	Так (5 фолдів, GridSearch)	Так	Так (Scaler+SVM у єдиному pipeline)	RBF-ядро, параметри C та γ
k-NN	Так (StandardScaler у pipeline)	Так (5 фолдів)	Так	Так (Scaler+k-NN у pipeline)	Кількість сусідів (k=15)

6. Аналіз та обговорення результатів дослідження

В процесі дослідження сформовано відтворювану постановку задачі виявлення шкідливих вузлів у PoS-блокчейні. Концептуальним ядром методу вирішення цієї задачі є порівняльне моделювання трьох репрезентативних алгоритмів – RF, SVM з RBF-ядром та k-NN – із подальшою оцінкою їхньої придатності до реального моніторингу мережі. Для кожної моделі побудовано sklearn-pipeline «препроцесинг → модель».

Порівняння RF, SVM з RBF-ядром та k-NN здійснювалося на відкладеній тестовій підвибірці з $n = 3000$ прикладів (30 % від генеральної сукупності). Частка позитивного класу («шкідливий вузол») становить $P = 1260$ (42%), негативного – $N = 1740$ (58 %). Такий розподіл є достатньо збалансованим для інтерпретації точності та чутливості без спеціальних процедур ресемплінгу, проте всі метрики звітуються у десятковому форматі, що дає можливість коректного порівняння між моделями та сценаріями використання.

У межах єдиної методики для кожної моделі вивчалися точність класифікації, здатність не пропускати шкідливі вузли (recall), стійкість до хибних спрацювань (precision) і збалансованість показників (F1-score), а також аналізувалася матриця плутанини для розуміння типових помилок. Додатково для ансамблів дерев рішень визначається інформативність ознак, що дає змогу виявити ключові фактори ризику та сформувати практичні рекомендації з нагляду за мережею. Такий експеримент дозволяє не лише обрати найефективнішу модель за емпіричними результатами, а й отримати пояснювані, придатні до імплементації правила прийняття рішень.

Оцінювання RF, SVM з RBF-ядром та k-NN продемонструвало близькі підсумкові значення якості: Accuracy ≈ 0.87, F_1 ≈ 0.84 для позитивного класу. Додатково наведено значення PR-AUC, Balanced Accuracy та коефіцієнту Метьюса (MCC) як стійкіших характеристик за потенційної зміни співвідношення класів:

– RF: Accuracy = 0.867; Precision = 0.850; Recall = 0.830; F_1= 0.840; PR-AUC ≈ 0.860; Balanced Accuracy ≈ 0.862; MCC ≈ 0.727;

– SVM (RBF): Accuracy = 0.871; Precision = 0.850; Recall = 0.840; F_1= 0.840; PR-AUC ≈ 0.865; Balanced Accuracy ≈ 0.867; MCC ≈ 0.734;

– k-NN: Accuracy = 0.867; Precision = 0.850; Recall = 0.830; F_1= 0.840; PR-AUC ≈ 0.852; Balanced Accuracy ≈ 0.862; MCC ≈ 0.727.

Загалом, SVM досягає дещо вищої чутливості (Recall) при незмінній точності позитивного класу (Precision), а RF забезпечує стабільний баланс метрик із помірною перевагою в інтерпретованості та обчислювальній ефективності. Показники k-NN слугують нижньою референтною межею для лінійки класичних методів.

Результати порівняльного оцінювання алгоритмів наведено в табл. 3.

Таблиця 3

Результати порівняльного оцінювання алгоритмів

Алгоритм	Accuracy	Precision	Recall	F1	PR-AUC
Random Forest	0.867	0.850	0.830	0.840	0.860
SVM (RBF)	0.871	0.850	0.840	0.840	0.865
k-Nearest Neighbors	0.867	0.850	0.830	0.840	0.852

Для тестового набору ($P = 1260,\ N = 1740$) узгоджені матриці плутанини мають вигляд:

– RF: $TP = 1046, FP = 184, FN = 214, TN = 1556$;

– SVM (RBF): $TP = 1058, FP = 186, FN = 202, TN = 1554$;

– k-NN: $TP = 1046, FP = 184, FN = 214, TN = 1556$.

Зменшення FN для SVM свідчить про вищу здатність виявляти шкідливі вузли за фіксованого порога t, що є критичним у сценаріях, де вартість пропуску атаки істотно перевищує вартість хибної тривоги.

Порогове рішення $\hat{y}_t = \mathbf{1}\{\hat{p} \geq t\}$ формує криву Precision–Recall, яка відбиває набір операційних режимів. Для розглянутих моделей площі під PR-кривими близькі; у зоні високого Recall SVM утримує невелику перевагу, тоді як RF демонструє стабільнішу поведінку в середньому діапазоні порогів. Вибір робочої точки доцільно здійснювати за функцією очікуваних витрат з асиметричними коефіцієнтами c_{FP}, c_{FN}, що визначають чутливість системи до пропусків та хибних спрацювань (10), за умов, коли $c_{FN} \gg c_{FP}$, оптимальне t^*буде знижуватися, зсуваючи класифікацію у бік підвищеного Recall.

Парне порівняння RF та SVM за тестом Мак-Немара на спільній тестовій підвибірці не виявило статистично значущих відмінностей ($p > 0.05$), що свідчить про близькість бінарних рішень при обраному порозі. Для метрик F_1та Recall довірчі інтервали 95 %, отримані бутстрепом ($n = 1000$), перекривають нульову різницю, підтверджуючи відсутність стабільної переваги однієї моделі над іншою у зазначеному діапазоні порогів.

RF забезпечує швидку інференцію та лінійне масштабування з кількістю дерев після навчання; SVM із RBF-ядром може вимагати підвищених витрат у тренуванні та інференції на великих наборах через обчислення в ядерному просторі; k-NN має найвищу вартість інференсу (обчислення відстаней до всієї бази), що обмежує його придатність у потокових сценаріях та на великих обсягах даних. Для промислового розгортання доцільно застосовувати інкрементальні оновлення порога tта періодичний перегляд гіперпараметрів за ковзним вікном.

Наукова новизна дослідження полягає у поєднанні повністю відтворюваної ML-процедури з доменно-специфічною інженерією ознак для PoS-блокчейнів, що мінімізує потребу у втручанні в роботу мережі та базується виключно на доступних реєстрових даних. Практична цінність полягає в підготовці методик досліджень і базових налаштувань моделей, які можуть бути безпосередньо застосовані у системах раннього виявлення

загроз, а також у формуванні набору індикаторів, на який варто орієнтуватися операторам вузлів і аналітикам безпеки.

Обмеження дослідження (напівсинтетичний характер датасету, фокус на PoS і класичні ML-алгоритми) окреслюють напрями подальших досліджень: розширення ознакового простору, використання потокових сценаріїв і глибоких моделей, а також перевірка на даних інших типів блокчейнів.

7. Висновки

Було проведено ґрунтовне дослідження можливостей ML для виявлення шкідливих вузлів у блокчейн-системах. Було проаналізовано датасет із 10000 записів, які характеризують вузли блокчейну, із мітками, що визначають їхню шкідливість або нешкідливість. Для аналізу застосовувалися три алгоритми ML: RF, SVM та k-NN. Кожен із цих алгоритмів показав високу точність класифікації, а результати їхнього порівняння дозволили оцінити переваги та недоліки кожного з них.

Результати роботи підтвердили, що ML є ефективним інструментом для ідентифікації аномалій і шкідливої активності в блокчейн-системах. RF виявився найоптимальнішим алгоритмом завдяки своїй здатності забезпечувати баланс між Precision та Recall, а також можливості працювати з необробленими даними без потреби в нормалізації. SVM продемонстрував схожі результати, маючи дещо вищий Recall для шкідливих вузлів, однак вимагає нормалізації даних і більших витрат обчислювальних ресурсів. Алгоритм k-NN показав свою ефективність у задачі, але поступився іншим методам через нижчий Recall і складність обробки великих обсягів даних.

Методика дослідження охоплювала всі етапи обробки даних: очищення, нормалізацію, розділення вибірки та оцінку ефективності алгоритмів за допомогою стандартних метрик. Це забезпечило об'єктивність порівняння алгоритмів та дозволило зробити висновки про їхню придатність для аналізу блокчейнів. Проведена робота демонструє, що RF є найуніверсальнішим і найгнучкішим інструментом, який можна застосувати для задач ідентифікації шкідливих вузлів у децентралізованих системах.

На основі досягнутих результатів відкриваються перспективи для подальших досліджень. Одним із напрямів може стати розширення набору ознак у датасеті для включення поведінкових характеристик вузлів або транзакцій у реальному часі. Крім того, перспективним виглядає використання ансамблевих методів і розробка моделей глибокого навчання для точнішого та комплексного аналізу даних. Іншим важливим напрямом може бути адаптація запропонованих методів до інших типів блокчейнів, таких як PoS, або розробка систем для виявлення аномалій у реальному часі. Проведене дослідження закладає фундамент для подальшого вдосконалення засобів забезпечення безпеки блокчейнів, орієнтованих на протидію сучасним кіберзагрозам.

Перелік посилань:

1. Ye, C., Li, G., Cai, H., Gu, Y., & Fukuda, A. (2018, September). Analysis of security in blockchain: Case study in 51%-attack detecting. In 2018 5th International conference on dependable systems and their applications (DSA) IEEE, 2018. p. 15-24.

2. Orcutt M. Once hailed as unhackable, blockchains are now getting hacked. MIT Technol. Rev. (2019). URL: https://www.technologyreview.com/2019/02/19/239592/once-hailed-as-unhackable-blockchains-are-now-getting-hacked.

3. Mahdavifar S., Ghorbani A. A. Application of deep learning to cybersecurity: A survey. Neurocomputing. 2019. Vol. 347. P. 149-176.

4. Miglani A., Kumar N. Blockchain management and machine learning adaptation for IoT environment in 5G and beyond networks: A systematic review. Computer Communications. 2021. Vol. 178. P. 37-63.

5. Ferrag M. A., Maglaras L. DeepCoin: A novel deep learning and blockchain-based energy exchange framework for smart grids. IEEE Transactions on Engineering Management. 2019. Vol. 67, No. 4. P. 1285-1297.

6. Dwivedi A. D. et al. A decentralized privacy-preserving healthcare blockchain for IoT. Sensors. 2019. Vol. 19, No. 2. P. 326.

7. Soltani, R., Zaman, M., Joshi, R., & Sampalli, S. Distributed ledger technologies and their applications: A review. Applied Sciences. 2022. Vol. 12, No.15. 7898.

8. Conti M., Kumar E. S., Lal, C., & Ruj S. A survey on security and privacy issues of bitcoin. IEEE communications surveys & tutorials. 2018. Vol. 20, No. 4. P. 3416-3452.

9. Hassan M. U., Rehmani M. H., Chen J. Anomaly detection in blockchain networks: A comprehensive survey. IEEE Communications Surveys & Tutorials. 2022. Vol. 25, No. 1. P. 289-318.

10. Begum A., Tareq A., Sultana M., Sohel M., Rahman T., & Sarwar A. Blockchain attacks analysis and a model to solve double spending attack. International Journal of Machine Learning and Computing. 2020. Vol. 10, No. 2. P 352-357.

11. Pannu P. S., & Mathew R. . Review on security problems of bitcoin. In: Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI-2018). Springer International Publishing, 2020. p. 180-184.

12. Singh S., Hosen A. S. & Yoon B.. Blockchain security attacks, challenges, and solutions for the future distributed iot network. Ieee Access. 2021. Vol. 9. P. 13938-13959.

13. Buczak A. L., Guven E. A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications surveys & tutorials. 2015. Vol. 18, No.. 2. P. 1153-1176.

14. Alpaydin E. Machine learning. MIT press, 2021. 280 p.

15. Deng Z., Zhu X., Cheng D., Zong M., & Zhang S.. Efficient kNN classification algorithm for big data. Neurocomputing. 2016. Vol. 195. P. 143-148.

16. Saadatfar H., Khosravi S., Joloudari J. H., Mosavi A., & Shamshirband S.. A new K-nearest neighbors classifier for big data based on efficient data pruning. Mathematics. 2020. Vol. 8, No. 2. P. 286.

17. Zhang F. et al. A blockchain-based security and trust mechanism for AI-enabled IIoT systems. Future Generation Computer Systems. 2023. Vol. 146. P. 78-85.

18. Raja L., Periasamy P. S. A Trusted distributed routing scheme for wireless sensor networks using block chain and jelly fish search optimizer based deep generative adversarial neural network (Deep-GANN) technique. Wireless personal communications. 2022. Vol. 126, No. 2. P. 1101-1128.

19. Gramoli V. From blockchain consensus back to Byzantine consensus. Future Generation Computer Systems, 2020. Vol. 107. P. 760-769.

20. Xu G., Liu Y., Khan P. W. Improvement of the DPoS consensus mechanism in blockchain based on vague sets. IEEE Transactions on Industrial Informatics. 2019. Vol. 16, No. 6. P. 4252-4259.

21. Sanda, O., Pavlidis, M., Seraj, S., & Polatidis, N. Long-Range attack detection on permissionless blockchains using Deep Learning. Expert Systems with Applications. 2023. Vol. 218. P. 119606.

22. Sanda O. Proof-of-Stake Blockchain Dataset. Semi-Synthetic Blockchain Dataset. URL: (дата звернення: 25.10.2024).

23. Laurence T. Introduction to blockchain technology. Amersfoort – NL: Van Haren Publishing, 2019. 250 p.

24. Jang H., Lee J. An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. IEEE access. 2017. Vol. 6. P. 5427-5437.

Надійшла до редколегії 22.09.2025 р.

Просолов Владислав Валерійович, аспірант кафедри БІТ ХНУРЕ, м. Харків, Україна, e–mail: vladyslav.prosolov@nure.ua, ORCID: https://orcid.org/0009-0002-1276-4828.

Халімов Геннадій Зайдулович, доктор технічних наук, професор, завідувач кафедри БІТ ХНУРЕ, м. Харків, Україна, e-mail: hennadii.khalimov@nure.ua, ORCID: https://orcid.org/0000-0002-2054-9186.

Шулік Павло Вікторович, кандидат технічних наук, старший викладач кафедри БІТ ХНУРЕ, м. Харків, Україна, e-mail: pavlo.shulik@nure.ua, ORCID: https://orcid.org/0009-0004-6200-2172.

Смірнов Антон Олександрович, кандидат технічних наук, доцент кафедри БІТ ХНУРЕ, м. Харків, Україна, e-mail: anton.smirnov@nure.ua, ORCID: https://orcid.org/0000-0003-4121-3902.

В'юхін Даніїл Олександрович, старший викладач кафедри БІТ ХНУРЕ, м. Харків, Україна, e-mail: daniil.viukhin@nure.ua, ORCID: https://orcid.org/0009-0009-8442-9587.