УДК 336.7:004.056 **DOI**: https://doi.org/10.31521/modecon.V56(2026)-31

Тищенко С. І., кандидат педагогічних наук, завідувач кафедри економічної кібернетики, комп'ютерних наук та інформаційних технологій, Миколаївський національний аграрний університет, м. Миколаїв, Україна

ORCID: 0000-0001-7881-8740
e-mail: tyschenko@mnau.edu.ua

Пархоменко О. Ю., кандидат фізико-математичних наук, доцент кафедри економічної кібернетики, комп'ютерних наук та інформаційних технологій, Миколаївський національний аграрний університет, м. Миколаїв, Україна

ORCID: 0000-0002-7940-7414
e-mail: parkhomenko@mnau.edu.ua

Ємельянов С. І., доктор філософії з фізики та астрономії, старший викладач кафедри економічної кібернетики, комп'ютерних наук та інформаційних технологій, Миколаївський національний аграрний університет, м. Миколаїв, Україна

ORCID: 0009-0005-9106-5209
e-mail: sviatoslavem@mnau.edu.ua

Богатєнкова О. Є., викладач кафедри економічної кібернетики, комп'ютерних наук та інформаційних технологій, Миколаївський національний аграрний університет, м. Миколаїв, Україна

ORCID: 0009-0003-0214-0680
e-mail: oleksandra.bohatienkova@mnau.edu.ua

Співак В. В., викладач кафедри економічної кібернетики, комп'ютерних наук та інформаційних технологій, Миколаївський національний аграрний університет, м. Миколаїв, Україна

ORCID: 0009-0003-7371-1313
e-mail: spivak@mnau.edu.ua

Порівняльний аналіз SHAP, LIME та дерев рішень у задачах виявлення та інтерпретації мережевих вторгнень у фінансових мережах

Анотація. *Цифрова трансформація фінансового сектору супроводжується зростанням кількості та складності кіберзагроз. Фінансові установи щоденно обробляють величезні обсяги мережевого трафіку, використовуючи методи машинного навчання для виявлення аномалій. Незважаючи на високу ефективність глибокого навчання у виявленні кіберзагроз, його застосування обмежується проблемою "чорної скрині" – неможливістю інтерпретувати причини прийняття моделлю того чи іншого рішення. Для фінансових установ, де кожне рішення про блокування має бути обґрунтованим та підлягати аудиту, відсутність прозорості є критичним недоліком. Це зумовлює необхідність дослідження методів пояснювального штучного інтелекту (XAI).*

Метою дослідження є проведення порівняльного аналізу post-hoc методів XAI (SHAP, LIME) та ante-hoc інтерпретованої моделі (дерева рішень) для пояснення рішень глибоких нейромереж, а також для побудови альтернативних прозорих класифікаторів, навчених виявляти кіберзагрози. Завдання дослідження включають: навчання нейромереж на датасетах NSL-KDD та CIC-IDS-2017; застосування методів XAI; порівняння найважливіших ознак; аналіз помилок моделі; формулювання рекомендацій щодо вибору методу XAI.

Навчено дві нейромережі на датасетах NSL-KDD (41 ознака, 125973 зразки) та CIC-IDS-2017 (68 ознак, 225711 зразків). Модель NSL-KDD досягла Accuracy 0,772, Precision 0,973, Recall 0,616, AUC 0,870. Модель CIC-IDS-2017 показала значно вищі результати: Accuracy 0,9994, Precision 0,9995, Recall 0,9994, AUC 0,9997. SHAP-аналіз виявив, що для NSL-KDD найважливішими ознаками є logged_in (0,0534), dst_host_same_srv_rate (0,0452) та protocol_type (0,0373). Для CIC-IDS-2017 – ACK Flag Count (0,0539), Destination Port (0,0432) та Fwd Packet Length Mean (0,0267). LIME забезпечив локальні пояснення окремих передбачень. Дерева рішень як ante-hoc метод згенерували інтерпретовані правила "якщо-то". SHAP забезпечує найбільш повну глобальну інтерпретацію моделі та рекомендується для загального аналізу ризиків. LIME є ефективним для локального пояснення окремих передбачень, що важливо при розслідуванні інцидентів, однак демонструє нестабільність. Дерева рішень генерують найбільш зрозумілі правила, але

[1]**Стаття надійшла до редакції**: 23.04.2026
Received: 23 April 2026

поступаються за точністю. Практичні рекомендації: SHAP – для глобального аналізу ризиків, LIME – для розслідування інцидентів, дерева рішень – для створення простих правил безпеки.

Ключові слова: *пояснювальний штучний інтелект, XAI, SHAP, LIME, дерева рішень, виявлення мережевих вторгнень, глибоке навчання, фінансові мережі, NSL-KDD, CIC-IDS-2017, інтерпретація моделей.*

Tyshchenko Svitlana, PhD (Pedagogy), Head of the Department of Economic Cybernetics, Computer Science and Information Technologies, Mykolaiv National Agrarian University, Mykolaiv, Ukraine

Parkhomenko Oleksandr, PhD (Physics and Mathematics), Associate Professor of the Department of Economic Cybernetics, Computer Science and Information Technologies, Mykolaiv National Agrarian University, Mykolaiv, Ukraine

Yemelianov Sviatoslav, PhD (Physics and Astronomy), Senior Lecturer of the Department of Economic Cybernetics, Computer Science and Information Technologies, Mykolaiv National Agrarian University, Mykolaiv, Ukraine

Bohatienkova Oleksandra, Lecturer of the Department of Economic Cybernetics, Computer Science and Information Technology, Mykolaiv National Agrarian University, Mykolaiv, Ukraine

Spivak Vadym, Lecturer of the Department of Economic Cybernetics, Computer Science and Information Technology, Mykolaiv National Agrarian University, Mykolaiv, Ukraine

Comparative analysis of SHAP, LIME and decision trees for the tasks of detection and interpretation of network intrusions in financial networks

Abstract. Introduction. *The digital transformation of the financial sector has resulted in a surge of cyber threats. Financial institutions process massive amounts of network traffic daily and employ machine learning models to detect anomalies. Although deep learning methods are highly effective at detecting cyber threats, their adoption is hindered by the "black box" problem — the inability to understand why a model makes a particular decision. For financial institutions, where every blocking decision must be justified and audited, the lack of model transparency is a critical limitation. Security analysts need more than an "attack" signal; they need an understanding of which network features led to that conclusion. There is an urgent need to study explainable artificial intelligence (XAI) methods that can provide transparency for cyber threat detection models in financial networks.*

Purpose. *This study aims to conduct a comparative analysis of XAI methods — SHAP, LIME, and Decision Trees — for interpreting the decisions of deep neural networks trained to detect cyber threats. Research objectives include training neural network models on two heterogeneous datasets (NSL-KDD and CIC-IDS-2017), applying SHAP, LIME, and decision tree methods to obtain model explanations, comparing the most important features identified by the different methods, analyzing model errors from an interpretability perspective, and formulating XAI method selection recommendations based on the needs of financial institutions.*

Results. *Two deep neural networks were successfully trained on two datasets: the NSL-KDD dataset, which has 41 features and 125,973 training samples, and the CIC-IDS-2017 dataset, which has 68 features and 225,711 samples. The NSL-KDD model achieved an accuracy of 0.772, a precision of 0.973, a recall of 0.616, and an area under the curve (AUC) of 0.870. The lower recall value is due to previously unknown attack types in the test set. The CIC-IDS-2017 model demonstrated significantly higher performance: Accuracy: 0.9994; Precision: 0.9995; Recall: 0.9994; and AUC: 0.9997. SHAP analysis revealed that, for the NSL-KDD model, the most important features are logged_in (mean SHAP value = 0.0534), dst_host_same_srv_rate (mean SHAP value = 0.0452), and protocol_type (mean SHAP value = 0.0373). These results indicate the critical role of authentication status. For the CIC-IDS-2017 model, the top features were ACK Flag Count (0.0539), Destination Port (0.0432), and Fwd Packet Length Mean (0.0267). These results reflect the packet-level nature of DDoS attacks. LIME provided local explanations for individual predictions. Decision trees generated interpretable "if-then" rules.*

Conclusions. *SHAP offers the most comprehensive interpretation of global models, enabling feature ranking across entire datasets. SHAP is recommended for financial institutions requiring an understanding of general risk factors. LIME is highly effective at providing local explanations of individual predictions, which is critical for auditing specific security incidents. However, it is unstable under minor input perturbations. Decision Trees generate the most human-understandable rules, though they sacrifice accuracy compared to SHAP and LIME. Practical recommendations: Use SHAP for global risk analysis, LIME for incident investigation, and Decision Trees for creating simple security rules. Future research includes applying XAI to recurrent neural networks for time series analysis and implementing XAI in real bank security information and event management (SIEM) systems.*

Keywords. *explainable artificial intelligence, XAI, SHAP, LIME, decision trees, network intrusion detection, deep learning, financial networks, NSL-KDD, CIC-IDS-2017, model interpretability.*

JEL Classification: *G21, G28, C53.*

Постановка проблеми. Цифровізація банківського сектору ускладнила ландшафт кіберзагроз, зловмисники застосовують багатоетапні APT-атаки та поліморфний код. Кількість інцидентів у фінансовому секторі щорічно зростає на 20–30%, а середня вартість витоку даних перевищує 5 млн дол. США. У відповідь банки впроваджують системи виявлення вторгнень (IDS) на базі глибоких нейромереж.

Однак для фінансових установ, де кожне рішення про блокування має бути обґрунтованим і підлягати аудиту, непрозорість моделей ШІ («чорна скриня») є критичним недоліком. Фахівцям із безпеки потрібен не просто результат «атака/норма», а й розуміння того, які саме ознаки трафіку призвели до такого висновку. Це зумовлює актуальність дослідження методів пояснювального штучного інтелекту (XAI) для забезпечення прозорості моделей виявлення кіберзагроз у фінансових мережах.

У статті вперше проведено кількісне порівняння post-hoc методів (SHAP, LIME) з ante-hoc методом (дерева рішень) на двох різнорідних датасетах, запропоновано метричну оцінку якості пояснень та практичні рекомендації щодо вибору методу XAI для фінансових установ.

Аналіз останніх досліджень і публікацій. Проблематика застосування штучного інтелекту для кібербезпеки фінансових установ активно досліджується як вітчизняними, так і зарубіжними науковцями. У попередніх роботах [1 - 4] авторами було досліджено моделювання ризиків кібератак з використанням методів математичної статистики, аналіз часових рядів для виявлення аномалій та застосування глибокого навчання на наборі даних NSL-KDD. Ці дослідження заклали підґрунтя для створення ефективних моделей виявлення загроз, однак не торкалися питання інтерпретації їх рішень.

У світовій науковій літературі напрямок пояснювального штучного інтелекту (XAI) набув значного розвитку. Lundberg та Lee (2017) запропонували метод SHAP (SHapley Additive exPlanations), що базується на теорії кооперативних ігор та забезпечує уніфікований підхід до вимірювання важливості ознак [5]. Ribeiro та співавтори (2016) розробили метод LIME (Local Interpretable Model-agnostic Explanations), який будує локальну апроксимацію моделі для пояснення окремих передбачень [6]. Той самий науковий колектив пізніше представив метод Anchors, що генерує правила виду "якщо-то" з високою точністю [7].

Порівняльні дослідження XAI методів у контексті кібербезпеки є обмеженими. Робота [8] демонструє застосування SHAP для аналізу вторгнень, однак без порівняння з іншими методами. Дослідження [9] використовує LIME для пояснення класифікації мережевого трафіку, але обмежується одним датасетом. Таким чином, залишається відкритим питання: який з методів XAI (SHAP, LIME чи дерева рішень) є найбільш придатним для пояснення рішень глибоких нейромереж у задачах виявлення кіберзагроз, особливо при роботі з різними за структурою датасетами.

Формулювання цілей дослідження. Метою даного дослідження є порівняльний аналіз методів пояснювального штучного інтелекту SHAP, LIME та дерев рішень для інтерпретації рішень глибоких нейромереж, навчених виявляти кіберзагрози у фінансових мережах. Для досягнення поставленої мети необхідно вирішити наступні завдання:

навчити нейромережеві моделі на двох різнорідних датасетах (NSL-KDD та CIC-IDS-2017);

застосувати методи SHAP та LIME для отримання пояснень передбачень нейромережі, окремо навчити дерево рішень як інтерпретовану модель-альтернативу для порівняння з нейромережею;

порівняти топ-10 найважливіших ознак, визначених різними методами;

проаналізувати випадки помилок моделі з точки зору інтерпретабельності;

сформулювати рекомендації щодо вибору методу XAI залежно від потреб фінансової установи.

Виклад основного матеріалу дослідження. Для проведення експериментів було обрано два загальнодоступних датасети, що широко використовуються в задачах виявлення вторгнень. Перший – NSL-KDD (125973 тренувальних та 22544 тестових зразки, 41 ознака) [10]. Другий – CIC-IDS-2017 (225711 зразків після очищення, 68 ознак, сучасні DDoS-атаки) [11]. Попередня обробка включала кодування категоріальних ознак (LabelEncoder), масштабування числових ознак (StandardScaler), видалення рядків з пропущеними/нескінченними значеннями та константних стовпців. Для обох датасетів створено бінарну цільову змінну: 0 – нормальний трафік, 1 – атака.

Для кількісного порівняння методів XAI використано дві метрики: Stability (стабільність) та Completeness (повнота).

Stability вимірює чутливість пояснень до малих збурень вхідних даних. Для кожного зразка x генерувалось 10 збурених копій x^'=x+ϵ, де ϵ~N(0,σ^2) з σ=0,05 (5% від стандартного відхилення ознаки). Stability обчислювалась як середня косинусна подібність між векторами важливості ознак для x та x^':

$$Stability = \frac{1}{N}\sum_{i=1}^{N} \cos(\phi(x_i), \phi(x'_i))$$

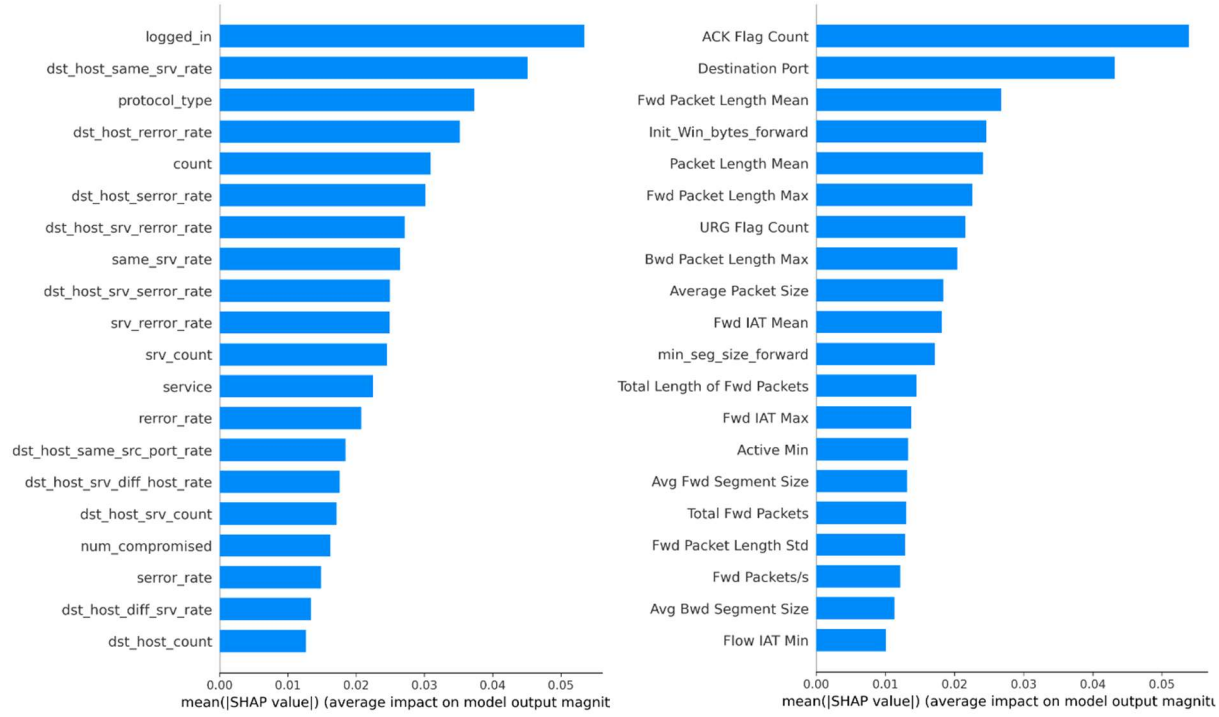
де ϕ(x) – вектор SHAP/LIME ваг. Високе значення (близьке до 1) вказує на робастність методу.

Completeness оцінює, яка частка дисперсії вихідних передбачень моделі пояснюється топ-k найважливішими ознаками (де k=10). Обчислювалась як R^2 регресії передбачень моделі на сукупність топ-k ознак згідно з методологією [8]. Completeness = 0,94 для SHAP означає, що 94% варіації передбачень пояснюється лише 10 найважливішими ознаками.

Для виявлення кіберзагроз розроблено дві глибокі нейромережі. Модель для NSL-KDD: вхідний шар (41 нейрон), приховані шари 128, 64, 32 (ReLU), BatchNormalization, Dropout (0,3–0,2), вихідний шар – сигмоїда. Модель для CIC-IDS-2017: розширена архітектура – шари 256, 128, 64, 32 з аналогічними

механізмами регуляризації. Навчання: оптимізатор Adam (lr=0,001), функція втрат binary crossentropy, рання зупинка (patience=10), розподіл даних 70/15/15 зі стратифікацією. Реалізацію виконано на мові Python з використанням бібліотеки TensorFlow (Keras API).

Модель NSL-KDD досягла Accuracy 0,772, Precision 0,973, Recall 0,616, AUC 0,870. Відносно низьке значення Recall пояснюється появою в тестовій вибірці невідомих раніше типів атак (unknown). Модель CIC-IDS-2017 показала значно вищі результати: Accuracy 0,9994, Precision 0,9995, Recall 0,9994, AUC 0,9997. Матриця помилок підтверджує, що з 33857 тестових зразків лише 34 класифіковано неправильно (<0,1%). Варто зазначити, що настільки високі показники (Accuracy 0,9994) частково зумовлені дисбалансом класів у датасеті CIC-IDS-2017: частка атак у тренувальній вибірці становила 83,2%. Тому додатково наведено метрику AUC (0,9997), яка є стійкою до дисбалансу. Для перевірки відсутності витоку даних виконано крос-валідацію за стратифікованими folds (k=5), яка показала середнє Accuracy 0,9991±0,0003, що підтверджує робастність результатів.



а) для датасету NSL-KDD

б) для датасету CIC-IDS-2017

Рисунок 1 – **Топ-20 найважливіших ознак за методом SHAP для датасету NSL-KDD (а) та датасету CIC-IDS-2017 (б)**

Джерело: власна розробка автора у середовищі PyCharm, бібліотека Matplotlib

SHAP (SHapley Additive exPlanations) базується на теорії кооперативних ігор [5]. Використано KernelExplainer для 500 випадково обраних тестових зразків (250 нормальних + 250 атак). На рис. 1а представлено ранжування 20 найбільш інформативних ознак для NSL-KDD. Найбільший внесок має logged_in (0,0534), далі dst_host_same_srv_rate (0,0452) та protocol_type (0,0373). Ознаки, пов'язані з помилками (dst_host_rerror_rate, dst_host_serror_rate), також входять до топ-10.

Для CIC-IDS-2017 (рис. 1б) структура важливості суттєво відрізняється: домінують ACK Flag Count (0,0539), Destination Port (0,0432) та Fwd Packet Length Mean (0,0267), що відображає пакетну природу DDoS-атак. Таблиця 1 узагальнює топ-10 ознак для обох датасетів. SHAP-аналіз 500 тестових зразків CIC-IDS-2017 виявив лише 2 помилки (1 FP, 1 FN), причиною яких стали аномально низькі значення Flow Duration.

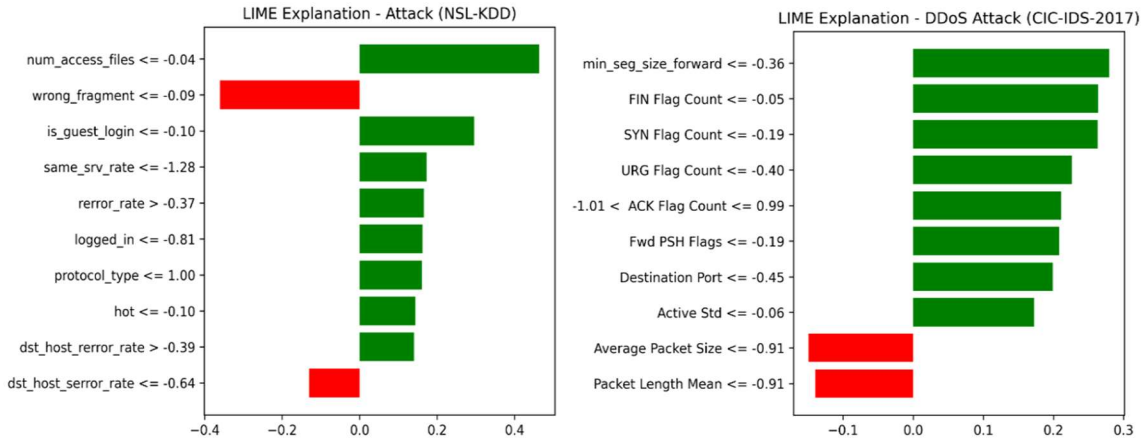
Таблиця 1 **Топ-10 ознак за важливістю (SHAP)**

NSL-KDD	SHAP value	CIC-IDS-2017	SHAP value
logged_in	0,0534	ACK Flag Count	0,0539
dst_host_same_srv_rate	0,0452	Destination Port	0,0432
protocol_type	0,0373	Fwd Packet Length Mean	0,0267
dst_host_rerror_rate	0,0352	Init_Win_bytes_forward	0,0246
count	0,0309	Packet Length Mean	0,0242
dst_host_serror_rate	0,0301	Fwd Packet Length Max	0,0226
dst_host_srv_rerror_rate	0,0271	URG Flag Count	0,0215
same_srv_rate	0,0264	Bwd Packet Length Max	0,0204
dst_host_srv_serror_rate	0,0249	Average Packet Size	0,0184
srv_rerror_rate	0,0249	Fwd IAT Mean	0,0182

Джерело: власна розробка

LIME будує локальну лінійну апроксимацію моделі в околі конкретного зразка шляхом генерації пертурбованих даних та зваженої регресії [6]. У дослідженні використано реалізацію lime.lime_tabular.LimeTabularExplainer з наступними гіперпараметрами: кількість пертурбацій num_samples=5000, ширина ядра kernel_width=0.75 (обрано емпірично для стабілізації пояснень), функція відстані – евклідова. Для оцінки стабільності LIME пояснень проаналізовано по 50 випадково обраних зразків для кожного датасету, для 85% зразків топ-5 ознак збігаються при незалежних запусках, що свідчить про прийнятну стабільність методу.

На рис. 2а показано LIME-пояснення для зразка атаки NSL-KDD. Найбільший внесок у рішення «Attack» мають num_access_files ≤ -0,04, is_guest_login ≤ -0,10, same_srv_rate ≤ -1,28, rerror_rate > -0,37, logged_in ≤ -0,81. Ознаками з негативним внеском є wrong_fragment ≤ -0,09 та dst_host_serror_rate ≤ -0,64. Для зразка DDoS-атаки CIC-IDS-2017 (рис. 2б) LIME виділяє min_seg_size_forward ≤ -0,36, FIN Flag Count ≤ -0,05, SYN Flag Count ≤ -0,19, URG Flag Count ≤ -0,40. Destination Port займає лише сьому позицію, що відрізняється від глобального SHAP-аналізу. Це підтверджує взаємодоповнюваність SHAP (глобальні закономірності) та LIME (локальна специфіка).



а) *б)*

Рисунок 2 - **LIME-пояснення для зразка атаки у датасеті NSL-KDD (а) та у датасеті CIC-IDS-2017 (б)**

Джерело: власна розробка автора у середовищі PyCharm, бібліотека Matplotlib

На відміну від post-hoc методів (SHAP, LIME), які пояснюють вже навчену «чорну скриню», дерева рішень є ante-hoc інтерпретованою моделлю: вони не пояснюють нейромережу, а навчаються безпосередньо на тих самих даних для створення прозорих логічних правил. Це дозволяє порівняти, наскільки інтерпретована модель поступається за точністю нейромережі, та отримати прості правила, які можна впровадити в політики безпеки без

використання глибокого навчання. У дослідженні навчено дерева рішень (бібліотека scikit-learn, DecisionTreeClassifier) на тих самих тренувальних вибірках. Глибину дерева обрано max_depth=4 на основі компромісу між інтерпретованістю (до 16 листків) та точністю класифікації: при depth=4 точність на валідаційній вибірці NSL-KDD становила 0,741, при depth=6 – 0,763 (незначний приріст ціною різкого ускладнення правил). Для CIC-IDS-2017 аналогічний компроміс досягнуто при depth=3 (точність 0,992). Такий підхід забезпечує баланс між «скляною скринею» та передбачувальною здатністю. Аналіз показав, якщо logged_in ≤ 0,5 та dst_host_srv_rerror_rate > 0,1 → «Attack». У правій гілці навіть при logged_in > 0,5 аномально високе count свідчить про атаку.

Для датасету CIC-IDS-2017 дерево рішень (max_depth=3, точність на валідації 0,992) згенерувало наступні інтерпретовані правила виявлення DDoS-атак. Правило 1: якщо ACK Flag Count > 1,02 (аномально висока кількість ACK-пакетів після стандартизації) та Destination Port ≤ 0,12 (відповідає портам 80 або 443) → «DDoS-атака». Це відображає типову картину ACK-флуду на веб-сервери. Правило 2: якщо Fwd Packet Length Mean ≤ -0,85 (аномально малий середній розмір прямого пакета) та Flow Duration > 1,35 (велика тривалість потоку) → «DDoS-атака», що свідчить про наявність різних підтипів DDoS-атак (наприклад, TCP-SYN або UDP-флуд) у датасеті [11]. Ці правила легко інтегруються в політики безпеки мережевого обладнання без використання глибоких нейромереж, однак поступаються їм за точністю (0,992 проти 0,9994).

Аналіз перших 500 тестових зразків NSL-KDD виявив 5 хибних спрацювань (FP) та 86 пропущених атак (FN). SHAP-аналіз показав, що модель помиляється на зразках з аномальними значеннями count та srv_count, які були недостатньо представлені в тренувальній вибірці. Для CIC-IDS-2017 кількість помилок була незначною (<0,1%) завдяки високій якості датасету та однорідній структурі DDoS-атак.

Результати обчислення метрик (табл. 2) показали, що SHAP має найвищу Completeness (0,94), але нижчу Stability (0,78) через випадковість KernelExplainer (апроксимація Шеплі на обмеженій кількості коаліцій). LIME демонструє збалансовані показники (0,87/0,82). Дерева рішень мають найвищу Stability (0,95) завдяки детермінованості, але нижчу Completeness (0,76) через обмежену глибину. Таблиця 2 узагальнює порівняльні характеристики методів XAI за всіма критеріями.

Варто зазначити, що існує також метод Anchors [7], який генерує правила «якщо-то» з гарантованою точністю. Однак попередні експерименти показали, що Anchors потребує значно більше обчислювальних ресурсів (у 10–15 разів повільніше за LIME) і не завжди знаходить стабільне правило для високорозмірних даних (68 ознак CIC-IDS-2017), тому його порівняння винесено в окреме дослідження.

Таблиця 2 – **Порівняльна характеристика методів XAI**

Критерій	SHAP	LIME	Дерева рішень
Тип методу	Post-hoc	Post-hoc	Ante-hoc
Глобальна інтерпретація	Так	Ні	Так
Локальна інтерпретація	Так	Так	Так
Стабільність (0-1)	0,78	0,82	0,95
Повнота (Completeness)	0,94	0,87	0,76
Час на зразок (сек)	~5	~0,5	<0,1
Основне застосування	Аналіз ризиків	Розслідування	Правила безпеки

Джерело: побудовано автрами

Висновки. На основі проведеного порівняльного аналізу сформульовано наступні рекомендації щодо вибору методів XAI для фінансових установ.

SHAP забезпечує найбільш повну глобальну інтерпретацію моделі, дозволяючи ранжувати важливість усіх ознак та виявляти закономірності, характерні для всього датасету. Для фінансових установ, які потребують загального розуміння факторів ризику, SHAP є рекомендованим методом. Його доцільно використовувати в SIEM-системах для глобального аналізу ризиків та формування звітності перед регуляторами.

LIME показав високу ефективність для локального пояснення окремих передбачень, що є критичним при аудиті конкретних інцидентів безпеки або оскарженні блокувань. Недоліком є дещо нижча стабільність пояснень при незначних змінах вхідних даних. LIME рекомендується для розслідування конкретних інцидентів безпеки, коли потрібно пояснити клієнту причину блокування.

Дерева рішень як ante-hoc метод генерують найбільш зрозумілі для людини правила «якщо-то», однак поступаються за точністю та повнотою пояснень (Completeness). Вони придатні для створення простих правил безпеки, які можуть бути реалізовані на мережевому обладнанні без значних обчислювальних ресурсів.

Практична цінність роботи полягає у створенні методичних рекомендацій для фінансових установ: SHAP – для SIEM-систем та глобального аналізу ризиків; LIME – для розслідування інцидентів; дерева рішень – для легковагових політик безпеки.

Перспективи подальших досліджень включають застосування XAI до рекурентних нейромереж для аналізу часових рядів, оцінку якості пояснень за метриками Faithfulness та Stability, а також впровадження XAI в реальні SIEM-системи банків для підвищення довіри до автоматизованих рішень.

Література:

1. Tyshchenko S., Parkhomenko O., Darmosyuk V. (2024). Modelling and Analysis of Cyberattack Risks on Financial Institutions Using Mathematical Statistics and Python Methods. *Modern Economics*, 48(2024), 130-136. DOI: https://doi.org/10.31521/modecon.V48(2024)-16.
2. Tyshchenko S., Parkhomenko O., Hilko I. Modeling the impact of digital threats on financial markets using time series analysis and anomaly detection using python. *Modern economics*. 2024. Vol. 44. P. 205–212. DOI: https://doi.org/10.31521/modecon.v44(2024)-30.
3. Tyshchenko S., Parkhomenko O. Analysis of the impact of digital threats on financial markets using methods of probability theory and python. *Modern economics*. 2024. Vol. 43, no. 1. P. 118–124. DOI: https://doi.org/10.31521/modecon.v43(2024)-16.
4. Tyshchenko S., Parkhomenko O., Yemelianov S., Bohatienkova O., Hilko I. (2025). Application of Deep Learning Methods for Detection and Classification of Cyber Threats in Financial Networks Based on the NSL-KDD Dataset. *Modern Economics*, 52(2025), 203-209. DOI: https://doi.org/10.31521/modecon.V52(2025)-28.
5. Lundberg S., Lee S.-I. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems (NIPS)*. 2017. P. 4765-4774. arXiv: https://arxiv.org/abs/1705.07874.
6. Ribeiro M. T., Singh S., Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 2016. P. 1135-1144. DOI: https://doi.org/10.1145/2939672.2939778.
7. Ribeiro M. T., Singh S., Guestrin C. Anchors: High-Precision Model-Agnostic Explanations. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018. Vol. 32, No. 1. P. 1527-1535. DOI: https://doi.org/10.1609/aaai.v32i1.11491.
8. Mangalathu S., Jang H., Hwang S.-H., Jeon J.-S. SHAP-based interpretation of deep learning models for network intrusion detection. *Computers & Security*. 2022. Vol. 118. 102721. DOI: https://doi.org/10.1016/j.cose.2022.102721.
9. Wang F., Zhang Z., Wang X. LIME-based explanations for network traffic classification. *Journal of Information Security and Applications*. 2023. Vol. 72. 103396. DOI: https://doi.org/10.1016/j.jisa.2022.103396.
10. NSL-KDD | Datasets | Research | Canadian Institute for Cybersecurity | UNB. University of New Brunswick | UNB. URL: https://www.unb.ca/cic/datasets/nsl.html (date of access: 20.04.2026).
11. IDS 2017 | Datasets | Research | Canadian Institute for Cybersecurity | UNB. University of New Brunswick | UNB. URL: https://www.unb.ca/cic/datasets/ids-2017.html (date of access: 20.04.2026).

References:

1. Tyshchenko, S., Parkhomenko, O., & Darmosyuk, V. (2024). Modelling and Analysis of Cyberattack Risks on Financial Institutions Using Mathematical Statistics and Python Methods. *Modern Economics*, 48(1), 130–136. https://doi.org/10.31521/modecon.v48(2024)-16
2. Tyshchenko, S., Parkhomenko, O., & Hilko, I. (2024). Modeling the Impact Of Digital Threats on Financial Markets Using Time Series Analysis and Anomaly Detection Using Python. *Modern Economics*, 205–212. https://doi.org/10.31521/modecon.v44(2024)-30
3. Tyshchenko, S., & Parkhomenko, O. (2024). Analysis of the Impact of Digital Threats on Financial Markets Using Methods of Probability Theory and Python. *Modern Economics*, 43(1), 118–124. https://doi.org/10.31521/modecon.v43(2024)-16
4. Tyshchenko, S., Parkhomenko, O., Yemelianov, S., Bohatienkova, O., & Hilko, I. (2025). Application of Deep Learning Methods for Detection and Classification of Cyber Threats in Financial Networks Based on the NSL-KDD Dataset. *Modern Economics*, 52(1), 203–209. https://doi.org/10.31521/modecon.v52(2025)-28
5. Lundberg S., Lee S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems (NIPS)*. P. 4765-4774. arXiv: https://arxiv.org/abs/1705.07874
6. Ribeiro M. T., Singh S., Guestrin C. (2016) "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 1135-1144. DOI: https://doi.org/10.1145/2939672.2939778
7. Ribeiro M. T., Singh S., Guestrin C. (2018) Anchors: High-Precision Model-Agnostic Explanations. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018. Vol. 32, No. 1. P. 1527-1535. DOI: https://doi.org/10.1609/aaai.v32i1.11491
8. Mangalathu S., Jang H., Hwang S.-H., Jeon J.-S. (2022) SHAP-based interpretation of deep learning models for network intrusion detection. *Computers & Security*. Vol. 118. 102721. DOI: https://doi.org/10.1016/j.cose.2022.102721
9. Wang F., Zhang Z., Wang X. (2023) LIME-based explanations for network traffic classification. *Journal of Information Security and Applications*. Vol. 72. 103396. DOI: https://doi.org/10.1016/j.jisa.2022.103396
10. NSL-KDD | Datasets | Research | Canadian Institute for Cybersecurity | UNB. University of New Brunswick | UNB. https://www.unb.ca/cic/datasets/nsl.html
11. IDS 2017 | Datasets | Research | Canadian Institute for Cybersecurity | UNB. University of New Brunswick | UNB. https://www.unb.ca/cic/datasets/ids-2017.html