УДК 336.7:004.056 **DOI**: https://doi.org/10.31521/modecon.V48(2024)-16

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

Дармосюк В. М., кандидат фізико-математичних наук, доцент кафедри вищої та прикладної математики, Миколаївський національний аграрний університет, м. Миколаїв, Україна

ORCID: 0000-0003-3275-8249
e-mail: darmosiuk@mnau.edu.ua

Моделювання та аналіз ризиків кібератак на фінансові установи з використанням методів математичної статистики та Python

Анотація. *У статті досліджено застосування методів математичної статистики та інструментів Python для моделювання та аналізу ризиків кібератак, з якими стикаються фінансові установи. Зростання масштабів і складності загроз кібербезпеки підкреслює вразливість цих установ, роблячи їх основними цілями для кіберзлочинців. Традиційні методи оцінки ризиків часто не справляються з адаптацією до еволюціонуючої природи цих загроз, що вимагає нових підходів, таких як автоматизований аналіз та прогнозне моделювання.*

У дослідженні продемонстровано, як математична статистика та Python можуть ідентифікувати ключові фактори ризику та прогнозувати потенційні атаки. Розроблено модуль на базі Python який інтегрує попередню обробку даних, кореляційний аналіз і моделювання машинного навчання для підвищення точності виявлення загроз. Різні прогностичні моделі, включаючи логістичну регресію, Random Forest і Gradient Boosting, було оцінено за допомогою наборів даних, таких як NSL-KDD, і виявлено високу точність у визначенні кіберзагроз. Запропоновані методи сприяють швидкому виявленню підозрілої активності, що покращує загальні заходи кібербезпеки. Майбутні напрямки досліджень включають інтеграцію методів глибокого навчання для аналізу складних патернів атак і адаптацію моделей до нових кіберзагроз. Підкреслено значення математичної статистики у розумінні кіберризиків, оскільки вона допомагає прогнозувати інциденти та оцінювати їхні наслідки. У зв'язку зі зростанням цифровізації фінансових послуг організації повинні пріоритетизувати надійні рамки кібербезпеки. Використовуючи Python і статистичні методи, фінансові установи можуть розробити ефективні стратегії для зменшення ризиків і забезпечення безпеки чутливих даних.

Експериментальним шляхом встановлено високу результативність використання програмного середовища Python як інструментального засобу для системного аналізу кіберризиків фінансових установ. Запропоновано підхід, який забезпечує автоматизований моніторинг, відносно швидке виявлення підозрілої активності та управління ризиками. Майбутні дослідження мають бути зосереджені на інтеграції глибокого навчання для аналізу складних патернів атак, адаптації моделей до нових кіберзагроз та розширенні джерел даних для покращення прогнозів.

Ключові слова *кібератаки; фінансові установи; математична статистика; Python; машинне навчання; аналіз ризиків; кореляційний аналіз; моделювання; кібербезпека; прогнозування.*

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

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

Darmosyuk Valentina, PhD (Physics and Mathematics), Associate Professor of the Department of Higher and Applied Mathematics, Mykolaiv National Agrarian University, Mykolaiv, Ukraine.

Modelling and Analysis of Cyberattack Risks on Financial Institutions Using Mathematical Statistics and Python Methods

[1]**Стаття надійшла до редакції**: 24.12.2024
Received: 24 December 2024

Abstract. Introduction *This article explores the use of mathematical statistical methods and Python tools to model and analyze the cyberattack risks faced by financial institutions. The growing scale and complexity of cybersecurity threats underscores the vulnerability of these institutions, making them prime targets for cybercriminals. Traditional risk assessment methods often fail to adapt to the evolving nature of these threats, requiring new approaches such as automated analysis and predictive modeling.*

Purpose. *The purpose of this research is to demonstrate the applicability of mathematical statistics and the Python language to analyze cyber risks, identify key risk factors, and predict attacks.*

Results. *The study demonstrates how mathematical statistics and Python can identify key risk factors and predict potential attacks. A Python-based module was developed that integrates data preprocessing, correlation analysis, and machine learning modeling to improve the accuracy of threat detection. Several predictive models, including logistic regression, random forest, and gradient boosting, were evaluated on datasets such as NSL-KDD and found to be highly accurate in identifying cyber threats. The results highlight the potential of Python as a powerful tool for automated monitoring and proactive risk management in financial institutions. The proposed methods contribute to the rapid detection of suspicious activity, which improves overall cybersecurity measures. Future research directions include the integration of deep learning methods to analyze complex attack patterns and to adapt models to new cyber threats. The importance of mathematical statistics in understanding cyber risks is emphasized, as it helps to predict incidents and assess their consequences. With the increasing digitization of financial services, organizations should prioritize a robust cybersecurity framework. By using Python and statistical methods, financial institutions can develop effective strategies to mitigate risk and ensure the security of sensitive data.*

Conclusions*. The results obtained highlight the potential of Python as a powerful tool for analysing cyber risks in financial institutions. An approach is proposed that enables automated monitoring, faster detection of suspicious activity, and risk management. Future research should focus on integrating deep learning to analyse complex attack patterns, adapting models to new cyber threats, and expanding data sources to improve predictions.*

Keywords *cyberattacks; financial institutions; mathematical statistics; Python; machine learning; risk analysis; correlation analysis; modeling; cybersecurity; forecasting.*

JEL Classification: *C63, C88.*

Постановка проблеми. Сучасний фінансовий сектор є однією з найпривабливіших цілей для кіберзлочинців, що зумовлено значною кількістю чутливих даних, великою фінансовою вигодою та ключовою роллю, яку фінансові установи відіграють в економічній стабільності. Згідно з останніми дослідженнями, кількість кібератак на банки, платіжні системи та страхові компанії постійно зростає. Наприклад, зловмисники використовують такі методи, як фішинг, DDoS-атаки, програми-вимагачі, атаки на банкомати та викрадення даних клієнтів. Втрати від таких інцидентів обчислюються мільярдами доларів і впливають не лише на конкретні установи, а й на всю фінансову систему. Наприклад, за даними сайту chainalysis.com, 2024 рік обіцяє стати найприбутковішим роком для програм-вимагачів, середній розмір викупу для найбільш небезпечних штамів з яких зріс з трохи менше 200000 доларів США на початку 2023 року до 1,5 мільйона доларів США в середині червня 2024 року [1].

Крім фінансових втрат, кібератаки підривають довіру клієнтів до банківської системи, що може мати довгострокові наслідки для економіки країни. Водночас кібербезпека залишається динамічною галуззю, яка вимагає нових підходів і технологій для аналізу ризиків.

Слід зауважити, що аналіз дозволяє не лише оцінити ймовірність інцидентів, але й виявити ключові фактори, що впливають на ризик. Це сприятиме створенню ефективних стратегій для мінімізації загроз та забезпечення більш надійного функціонування фінансової системи.

Зростання цифровізації фінансових послуг, включаючи онлайн-банкінг, мобільні платежі та блокчейн, значно підвищує залежність від інформаційних технологій. Разом із цим посилюються ризики, пов'язані з кіберзагрозами. За даними глобальних звітів, близько 70% фінансових установ зазнають принаймні однієї кібератаки на рік, і ця тенденція зростає.

Традиційні методи аналізу ризиків не завжди враховують динамічний характер кібератак і не здатні оперативно адаптуватися до нових методів зловмисників. Використання математичної статистики у поєднанні з інструментами Python дозволяє вирішувати ці проблеми завдяки автоматизованому аналізу великих обсягів даних у реальному часі, можливості налаштування моделей під специфічні потреби організацій та використанню статистичних методів для прогнозування ризиків.

Python є однією з найпопулярніших мов програмування в галузі аналізу даних, що обумовлено широким спектром бібліотек (Pandas, NumPy, Scikit-learn, Seaborn), які дозволяють ефективно працювати з даними, будувати моделі та візуалізувати результати.

Таким чином, використання методів математичної статистики та Python відкриває нові перспективи для аналізу та моделювання ризиків кібератак у фінансовому секторі. **Аналіз останніх досліджень та публікацій**. Зростаючі загрози кібербезпеки стають предметом активних досліджень, що підтверджується численними науковими роботами.

Metcalf L.та Casey W. досліджують математичні концепції для практики кібербезпеки [2]. Автори акцентують увагу на методах аналізу та візуалізації даних, охоплюючи ключові теми, такі як теорія графів та стійка гомологія. Python

використовується для аналізу даних і побудови моделей, які включають статистичні підходи.

Дослідження, проведене M. L. Sanni та ін., стосується прогнозування кіберзагроз для служб мобільних грошей [3]. Вони пропонують модель, що використовує Python для симуляцій та аналізу ризиків, наприклад, у контексті прогнозування атак на системи мобільних фінансів та машинне навчання для виявлення підозрілих клієнтів, що покращує загальну безпеку мобільних фінансових послуг.

C. Ieracitano та ін. пропонують систему глибокого навчання для виявлення вторгнень, яка використовує статистичний аналіз [4]. Ця модель демонструє значний потенціал у поліпшенні виявлення загроз у порівнянні з існуючими методами. У цьому дослідженні Python використовується для розробки глибокої автоенкодерної моделі виявлення вторгнень із використанням NSL-KDD набору даних. Модель покращує точність виявлення загроз за допомогою візуалізації великих даних і статистичного аналізу.

Zimba A. представляє байєсівський підхід до моделювання атак на фінансові установи за допомогою злочинного програмного забезпечення [5]. Методологія забезпечує детальне уявлення про структури атак і можливості їх пом'якшення. У цій роботі Python використовується для моделювання мережевих атак у фінансовому секторі з використанням байєсових мереж.

Дослідження Giudici P. та Raffinetti E. розглядає пояснювальні AI-моделі для управління кіберризиками з використанням Shapley values, які інтегруються зі статистичними методами [6]. Python використовується для побудови моделей і обчислення ризиків у фінансових установах, забезпечуючи підвищену прозорість і точність ризикових прогнозів.

У доповіді A. V. Alegria та ін. пропонується метод кількісного аналізу ризиків із застосуванням Python для моделювання ризиків у різних шарах фінансових систем, таких як шари презентації, бізнес-логіки та управління даними [7].

Tyshchenko S., Parkhomenko O. дослідили вплив цифрових загроз, включаючи кібератаки та крадіжки даних, на фінансові ринки, використовуючи теорію ймовірностей та програмування на Python для оцінки ризиків і розробки стратегій [8]. Проаналізовано використання байєсівських мереж та методів Монте-Карло для обчислення ймовірності успішних кібератак і застосовування моделей ARIMA для фінансового прогнозування, продемонстровано інтеграцію цих технік для підвищення стабільності ринків в умовах зростаючих кіберризиків.

Tyshchenko S., Parkhomenko O., Hilko I. розробили методологію моделювання та аналізу впливу кіберзагроз на фінансові ринки шляхом інтеграції методів аналізу часових рядів, алгоритмів виявлення аномалій та реалізації на мові Python [9].

Ці дослідження підкреслюють різноманітність підходів до моделювання та управління ризиками в сфері кібербезпеки, вказуючи на важливість інноваційних рішень для забезпечення фінансової безпеки.

Формулювання цілей дослідження. Дослідження має на меті продемонструвати, як можна застосовувати статистичні методи та програмування на Python для аналізу й моделювання ризиків, пов'язаних з кібератаками на фінансовий сектор.

Виходячи з мети дослідження, основні завдання можна сформулювати наступним чином:

1. Проаналізувати сучасні тенденції та масштаби кібератак на фінансові установи та визначити основні типи атак, які найчастіше спричиняють значні фінансові втрати.

2. Розробити програмний модуль на Python для автоматизованого аналізу даних та моделювання ризиків.

3. Реалізувати та оцінити прогностичні моделі машинного навчання для виявлення та прогнозування ризиків.

4. На основі отриманих результатів сформулювати рекомендації щодо того, як математична статистика та програмні інструменти можуть бути використані для управління кіберризиками.

Виклад основного матеріалу дослідження. Фінансові установи є основними цілями кібератак через доступ до конфіденційних даних і можливість отримання фінансових вигод. Дослідження показують, що кібератаки на ці установи можуть проявлятися в різних формах, зокрема фішингові атаки, які складають близько 40% усіх загроз [10, 11]. Зловмисники використовують соціальну інженерію для отримання паролів і кредитних карток. За даними SlashNext у 2023 році кількість шкідливих фішингових листів зросла на 1265% [12].

Іншою серйозною загрозою є DDoS-атаки, які перевантажують сервери банків, унаслідок чого клієнти не можуть отримати доступ до сервісів. У першій половині 2024 року кількість DDoS-атак зросла на 46%, досягнувши 445000 [13]. Атаки програм-вимагачів також є серйозною загрозою, з рекордним викупом понад 1 мільярд доларів у 2023 році, включаючи платіж у розмірі 15 мільйонів доларів від Caesars [1].

Атаки не лише завдають фінансових втрат, але й підривають довіру клієнтів, тому фінансовим установам важливо впроваджувати заходи безпеки.

Математична статистика є важливим інструментом для виявлення закономірностей і оцінки ризиків у кібербезпеці. Вона допомагає організаціям прогнозувати ймовірність атак, виявляти аномалії та оцінювати наслідки

кібератак. Використовуючи статистичні моделі та техніки, як регресійний аналіз і класифікаційні моделі, організації можуть краще підготуватися до потенційних загроз. Байєсові моделі дозволяють адаптувати ймовірності атак на основі нових даних, підтримуючи ефективну безпеку. Статистичний аналіз також допомагає оцінювати фінансові втрати і планувати бюджети на випадок кібератак.

Python є потужним інструментом для аналізу даних у кібербезпеці завдяки своїм бібліотекам, таким як Pandas та NumPy для обробки даних, а також Scikit-learn для машинного навчання. Цей інструмент дозволяє проводити швидкий аналіз журналів безпеки та будувати прогностичні моделі, що суттєво скорочує час реагування на загрози.

Для моделювання та аналізу ризиків кібератак на фінансові установи розроблено програмний модуль на Python, що реалізує автоматизовану обробку даних, статистичний аналіз та моделювання за допомогою методів машинного навчання. Код забезпечує інтеграцію всіх необхідних етапів аналізу від попередньої обробки даних до навчання моделей та оцінки їхньої точності.

Програмний модуль працює з двома файлами даних: KDDTrain+.csv (навчальна вибірка) та KDDTest+.csv (тестова вибірка), які є частиною широко відомого набору NSL-KDD [14]. Цей набір використовується для аналізу мережевих атак і систем виявлення вторгнень та включає наступні етапи реалізації алгоритму.

1. Завантаження та підготовка даних. На першому етапі програма завантажує дані з файлів KDDTrain+.csv та KDDTest+.csv. У процесі завантаження виконується перевірка наявності заголовків колонок і, за необхідності, їх автоматичне встановлення. Якщо заголовки відсутні або не відповідають очікуваним, програма додає стандартні назви колонок. Також перевіряється відповідність колонок між навчальним та тестовим наборами. Відсутні колонки у тестовому наборі додаються зі значеннями NaN, щоб забезпечити сумісність даних під час трансформації та аналізу. Для кожного набору даних визначаються два типи ознак: числові (numeric_features) ознаки, що мають тип даних int або float та категоріальні (categorical_features) текстові ознаки, що мають тип object. Якщо у даних відсутній стовпець attack_type, що є цільовою змінною для аналізу, програма створює його та заповнює значенням normal. Це гарантує коректну роботу подальших етапів.

2. Попередня обробка даних. Цей етап є критично важливим у підготовці наборів даних для аналізу та моделювання. У цьому процесі використовується модуль Column Transformer, який виконує кілька важливих операцій, що підвищують якість даних. Для числових ознак одним із перших кроків є обробка пропущених значень. Це досягається заповненням цих прогалин медіанним значенням відповідної ознаки, що допомагає зберегти цілісність даних. Після цього числові дані підлягають масштабуванню, щоб відповідати стандартному нормальному розподілу, який характеризується середнім значенням 0 і стандартним відхиленням 1. Ця нормалізація є важливою, оскільки забезпечує рівний внесок усіх числових ознак у ефективність моделі, запобігаючи нерівномірному впливу жодної окремої ознаки на результати. На відміну від цього, категоріальні ознаки обробляються інакше. Пропущені значення в цих ознаках заповнюються значенням "missing", що дозволяє моделі враховувати ці пробіли, не втрачаючи цілісності набору даних. Крім того, категоріальні текстові змінні кодуються за допомогою техніки OneHotEncoder. Цей метод створює бінарні представлення категорій, перетворюючи якісні дані в формат, який може бути ефективно використаний алгоритмами машинного навчання. Оброблені дані підсумовуються у формуванні матриць ознак, зокрема X_train, X_test. Ці матриці є важливими для навчання та тестування моделей. Завдяки ретельній попередній обробці дані стають більш придатними для наступного етапу моделювання, що врешті-решт призводить до покращення результатів у завданнях аналізу даних.

3. Аналіз даних включає в себе кілька ключових аспектів. Перш за все, на стадії описової статистики обчислюються основні статистичні характеристики як числових, так і категоріальних ознак. Серед таких характеристик можна виділити середнє значення, стандартне відхилення, мінімум і максимум для числових змінних. Для текстових змінних важливим є також визначення частоти появи різних категорій. Далі, при виконанні кореляційного аналізу, будується матриця кореляцій, яка ілюструє силу взаємозв'язку між різними змінними. Ці кореляції візуалізуються у вигляді теплової карти (heatmap), яка зберігається у графічному файлі під назвою correlation_matrix.png. Для зручності інтерпретації значення кореляції округлюються до двох знаків після коми. Цей етап аналізу допомагає краще зрозуміти, як змінні взаємодіють одна з одною, і виявити потенційні закономірності у даних.

4. Навчання моделей та оцінка ефективності. Для прогнозування ризиків кібератак використовуються три моделі машинного навчання. Першою моделлю є логістична регресія, яка забезпечує базову лінійну класифікацію. Ця модель слугує основним підходом до розуміння взаємозв'язків між вхідними ознаками та ймовірністю виникнення кібератаки. Другою моделлю є Random Forest, ансамблева модель, що складається з дерев рішень. Ця модель є особливо цінною, оскільки враховує нелінійні

взаємозв'язки між ознаками, що дозволяє більш детально зрозуміти дані. Об'єднуючи кілька дерев, Random Forest підвищує точність прогнозів і зменшує ризик перенавчання. Третьою моделлю є Gradient Boosting, яка використовує послідовне навчання дерев рішень для підвищення точності класифікації. Цей алгоритм будує дерева одне за одним, причому кожне нове дерево має на меті виправити помилки, допущені попередніми. Цей ітеративний підхід дозволяє Gradient Boosting досягати високої продуктивності на складних наборах даних.

Моделі тренуються на навчальному наборі, що складається з ознак і міток, представлених як (X_train, y_train), і потім тестуються на окремому тестовому наборі, позначеному як (X_test, y_test). Оцінка ефективності цих моделей є критично важливою і включає кілька ключових метрик. Classification Report надає важливі метрики, такі як точність, повнота та F1-міра для кожного класу, що допомагає всебічно оцінити продуктивність моделі.

Крім того, для багатокласової класифікації використовується метрика AUC-ROC (площа під кривою), яка надає уявлення про здатність моделі розрізняти різні класи. Нарешті, матриця плутанини слугує візуальним представленням правильних і неправильних прогнозів моделі, що полегшує зрозуміння того, де модель досягає успіху, а де може вимагати поліпшення.

Завдяки цьому структурованому підходу до навчання та оцінки моделей машинного навчання ми можемо краще прогнозувати та зменшувати ризики, пов'язані з кібератаками, що в кінцевому підсумку підвищує заходи кібербезпеки.

5. Візуалізація результатів. Всі графіки та результати обчислень зберігаються у вигляді файлів, що забезпечує легкий доступ до них для подальшого аналізу. Одним з ключових інструментів у аналізі ризиків кібератак для фінансових установ є матриця кореляцій, яка ілюструє силу взаємозв'язку між різними змінними. Ці кореляції візуалізуються у вигляді теплової карти (heatmap) (рисунок 1).

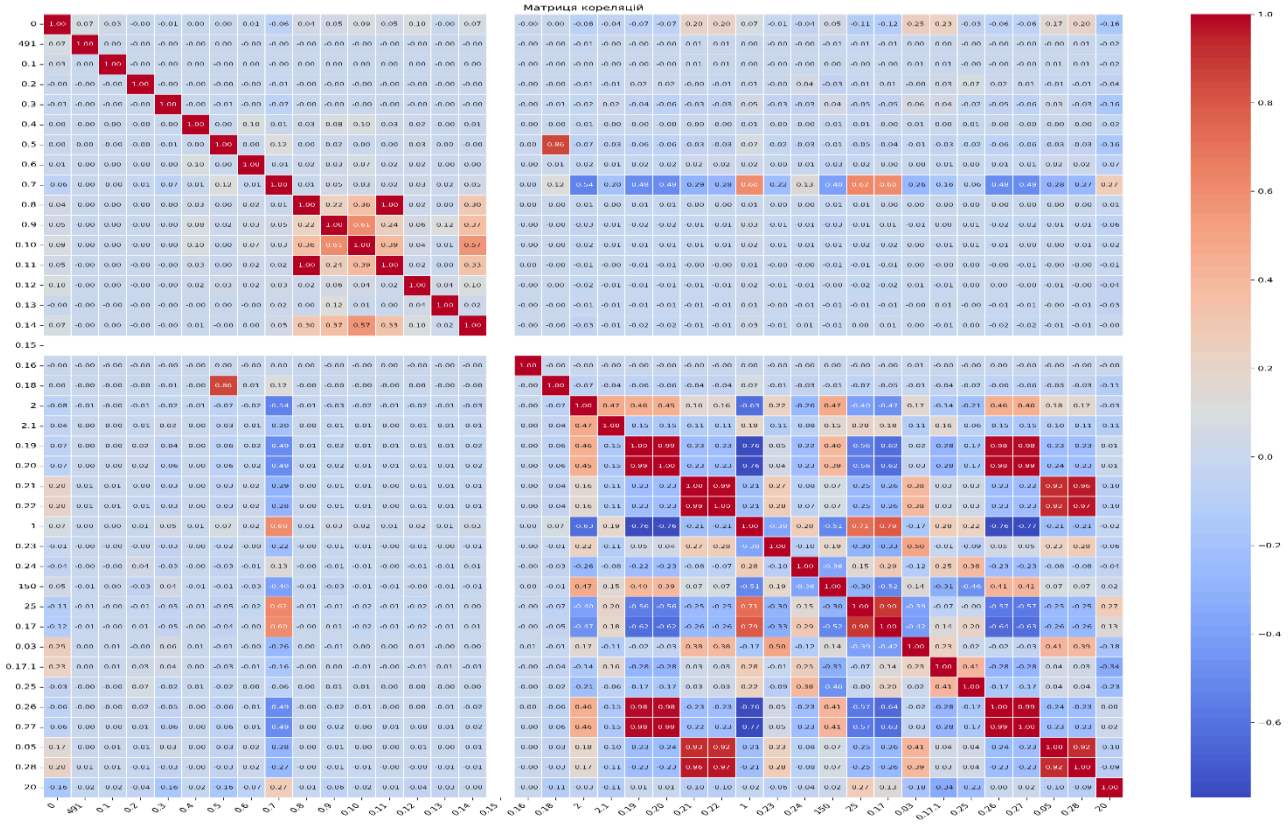

Рисунок 1 – **Матриця кореляцій у вигляді теплової карти (heatmap)**

Джерело: власна розробка авторів

Матриця кореляцій є важливим інструментом для моделювання та аналізу ризиків кібератак на фінансові установи. Вона дозволяє ідентифікувати ключові змінні, які значно впливають на класифікацію типів атак. Високий ступінь кореляції між ознакою та цільовою

змінною, такою як тип атаки (attack_type), вказує на її важливість у процесі класифікації. Наприклад, змінні, що характеризують частоту помилок чи аномалій у трафіку, можуть слугувати критичними індикаторами потенційних ризиків. Крім того, матриця кореляцій допомагає зменшити надмірність даних, виявляючи взаємно корельовані ознаки, які не додають нової інформації. Ця надмірність може ускладнити моделювання та призвести до "перенавчання", коли моделі добре працюють на навчальних даних, але не можуть узагальнити нову інформацію. Аналізуючи матрицю кореляцій, можна виявити та видалити зайві змінні, що спростить набір даних для більш ефективного моделювання. Розуміння взаємозв'язків між різними ознаками є ще однією значною перевагою використання матриці кореляцій. Наприклад, зв'язок між певними характеристиками мережевого трафіку, такими як кількість помилок і швидкість передачі даних, може поглибити розуміння механізмів атак. Це глибше розуміння є важливим для розробки більш точних моделей прогнозування ризиків.

Врешті-решт, завдяки аналізу матриці кореляцій можна визначити, які фактори найбільше впливають на ризики. Ці знання дозволяють оптимізувати моделі, що веде до більш ефективного виявлення кібератак. Таким чином, матриця кореляцій грає центральну роль у підвищенні безпеки та стійкості фінансових установ до потенційних кіберзагроз.

Запропонований підхід аналізу даних за допомогою математичної статистики та моделей машинного навчання дозволяє фінансовим установам виявляти ризики кібератак. Наприклад, кореляційний аналіз може виявити ключові фактори, що впливають на ймовірність атак, а моделі прогнозування допомагають класифікувати типи загроз у реальному часі. Це дає змогу організаціям краще усвідомлювати потенційні ризики та вчасно вживати заходів для їх мінімізації. Завдяки цьому підходу установи можуть впроваджувати системи автоматизованого моніторингу, які аналізують трафік і виявляють підозрілу активність. Автоматизація моніторингу значно скорочує час реакції на інциденти, що, в свою чергу, мінімізує втрати та захищає критичну інфраструктуру. Таким чином, організації отримують можливість своєчасно реагувати на загрози, забезпечуючи безпеку своїх систем. Крім того, методи машинного навчання, використані в моделюванні, дозволяють передбачати майбутні сценарії атак на основі історичних даних. Це сприяє проактивному управлінню ризиками і підвищенню стійкості до нових загроз. Завдяки такому прогнозуванню, фінансові установи можуть не лише реагувати на поточні виклики, але й готуватися до можливих майбутніх атак, що робить їх більш адаптивними у швидко змінюваному світі кібербезпеки.

Слід зазначити, що ефективність моделей даних значною мірою залежить від якості та повноти вхідних даних. Коли дані відсутні або некоректні, це може призвести до зниження точності прогнозів. Ця залежність від високоякісних даних підкреслює важливість ретельного збору та валідації даних. Крім того, одна з проблем, з якою стикаються ці моделі, полягає в їх непередбачуваності перед новими типами атак. Оскільки ландшафт загроз еволюціонує, моделі можуть мати труднощі в виявленні нових методів атак, які раніше не були представлені в навчальних вибірках. Це обмеження підкреслює необхідність постійного оновлення та адаптації моделей до нових загроз. Додатково, існує проблема високої обчислювальної складності, пов'язана з обробкою великих обсягів даних і навчанням моделей. Ця вимога до значних ресурсів може створити труднощі для організацій, особливо тих, які мають обмежену обчислювальну потужність. Таким чином, хоча моделі даних можуть надати цінні інсайти, їх ефективна реалізація вимагає ретельного врахування цих можливих недоліків.

Подальші дослідження можуть зосередитися на інтеграції додаткових джерел даних, таких як реальний мережевий трафік або лог-файли систем, для підвищення точності моделей. Перспективним напрямком є розробка систем адаптивного машинного навчання, що зможуть автоматично враховувати нові типи атак. Іншим важливим аспектом є впровадження методів глибокого навчання для аналізу складних патернів атак та їхньої ідентифікації.

Запропонований програмний код демонструє, як сучасні інструменти Python можна використовувати для аналізу та моделювання ризиків кібератак. Його реалізація показує високу ефективність підходу завдяки інтеграції статистичних методів і машинного навчання. Цей підхід може бути адаптований для аналізу ризиків в інших сферах, що підтверджує його універсальність.

Висновок. У статті досліджено можливості використання методів математичної статистики та Python для моделювання та аналізу ризиків кібератак на фінансові установи. Запропонований підхід дозволяє інтегрувати автоматизовану обробку даних, статистичний аналіз та моделювання, що значно підвищує ефективність і точність оцінки ризиків.

Особливу увагу приділено використанню Python як інструменту для аналізу великих обсягів даних і побудови моделей машинного навчання. Реалізований програмний модуль забезпечує ідентифікацію ключових змінних, прогнозування ризиків та оптимізацію систем захисту. Побудова матриці кореляцій дозволила визначити взаємозв'язки між ознаками, а використання моделей машинного навчання – підвищити точність виявлення загроз.

Результати дослідження можуть бути застосовані для створення систем автоматизованого моніторингу, які здатні оперативно реагувати на підозрілу активність та прогнозувати майбутні атаки. Однак ефективність реалізації залежить від якості даних і регулярного оновлення моделей, щоб враховувати нові загрози.

Подальші дослідження можуть бути спрямовані на інтеграцію глибокого навчання для аналізу складних патернів атак, адаптацію моделей до нових викликів у сфері кібербезпеки та розширення джерел даних. Запропонований підхід підтвердив свою універсальність і перспективність для вирішення завдань кібербезпеки у фінансовому секторі.

Література:

1. Chainalysis Team. 2024 crypto crime mid-year update part 1: cybercrime climbs. *Chainalysis*. 2024. 15 серпня. URL: https://www.chainalysis.com/blog/2024-crypto-crime-mid-year-update-part-1.
2. Metcalf L., Casey W. Cybersecurity and applied mathematics. Elsevier, 2016. 190 p. URL: https://doi.org/10.1016/c2015-0-01807-x.
3. A predictive cyber threat model for mobile money services / M. L. Sanni et al. *Annals of emerging technologies in computing*. 2023. Vol. 7, no. 1. P. 40–60. URL: https://doi.org/10.33166/aetic.2023.01.004.
4. Statistical analysis driven optimized deep learning system for intrusion detection / C. Ieracitano et al. *Advances in brain inspired cognitive systems*. Cham, 2018. P. 759–769. URL: https://doi.org/10.1007/978-3-030-00563-4_74.
5. Zimba A. A bayesian attack-network modeling approach to mitigating malware-based banking cyberattacks. *International journal of computer network and information security*. 2021. Vol. 14, no. 1. P. 25–39. URL: https://doi.org/10.5815/ijcnis.2022.01.03.
6. Giudici P., Raffinetti E. Explainable AI methods in cyber risk management. *Quality and reliability engineering international*. 2021. Vol. 38, no. 3. P. 1318-1326. URL: https://doi.org/10.1002/qre.2939.
7. Method of quantitative analysis of cybersecurity risks focused on data security in financial institutions / A. V. Alegria et al. *2022 17th iberian conference on information systems and technologies (CISTI)* (Madrid, Spain, 22–25 June 2022). 2022. P. 1-7. URL: https://doi.org/10.23919/cisti54924.2022.9820198.
8. 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. URL: https://doi.org/10.31521/modecon.v43(2024)-16.
9. 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. URL: https://doi.org/10.31521/modecon.v44(2024)-30.
10. Тенденції у розвитку фішингу та протидія йому. *IT Specialist*. 2024. 10 червня. URL: https://my-itspecialist.com/trends-in-phishing-development-and-countermeasures.
11. Різке зростання фішингових атак на фінансовий сектор: нові виклики кібербезпеки у 2024 році. *CyberSecureFox*. 2024. 15 листопада.URL: https://cybersecurefox.com/uk/zrostannya-fishingovyh-atak-finansovyi-sektor-2024/.
12. 12.The state of phishing 2023. Pleasanton CA : SlashNext security products, 2024. 28 p. URL: https://slashnext.com/wp-content/uploads/2023/10/SlashNext-The-State-of-Phishing-Report-2023.pdf.
13. Guards. Зростання кількості DDoS-атак на 46% у першому півріччі 2024 року. *Cybersecurity Services (Advisory/Ethical Hacking)*. URL: https://10guards.com/ua/blog/2024/09/23/surge-in-ddos-attacks-gcore-report-reveals-46-increase-in-first-half-of-2024.
14. GitHub - Jehuty4949/NSL_KDD: NSL-KDD Dataset. *GitHub*. 2015. URL: https://github.com/Jehuty4949/NSL_KDD.

References:

1. Chainalysis. (2024, July 18). 2024 Crypto Crime Mid-Year Update: Part 1. *Chainalysis Blog*. https://www.chainalysis.com/blog/2024-crypto-crime-mid-year-update-part-1.
2. Metcalf, L., & Casey, W. (2016). *Cybersecurity and applied mathematics*. Elsevier. 190 p. https://doi.org/10.1016/c2015-0-01807-x..
3. Sanni, M. L., Akinyemi, B. O., Olalere, D. A., Olajubu, E. A., & Aderounmu, G. A. (2023). A Predictive Cyber Threat Model for Mobile Money Services. *Annals of Emerging Technologies in Computing*, 7(1), 40–60. https://doi.org/10.33166/aetic.2023.01.004.
4. Ieracitano, C., Vinci, A., Adeel, A., Morabito, F. C., & Hussain, A. (2018). Statistical analysis driven optimized deep learning system for intrusion detection. In *Advances in Brain Inspired Cognitive Systems* (pp. 759–769). Cham: Springer. https://doi.org/10.1007/978-3-030-00563-4_74.
5. Zimba, A. (2021). A Bayesian Attack-Network Modeling Approach to Mitigating Malware-Based Banking Cyberattacks. *International Journal of Computer Network and Information Security*, 14(1), 25–39. https://doi.org/10.5815/ijcnis.2022.01.03.
6. Giudici, P., & Raffinetti, E. (2021). Explainable AI methods in cyber risk management. *Quality and Reliability Engineering International*, 38(3), 1318–1326. https://doi.org/10.1002/qre.2939.
7. Alegria, A. V., Loayza, J. L. M., Montoya, A. N., & Armas-Aguirre, J. (2022). Method of Quantitative Analysis of Cybersecurity Risks Focused on Data Security in Financial Institutions. *2022 17th Iberian Conference on Information Systems and Technologies (CISTI). IEEE*. https://doi.org/10.23919/cisti54924.2022.9820198.

8. 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.
9. 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*, 44, 205–212. https://doi.org/10.31521/modecon.v44(2024)-30.
10. IT Specialist. (2024, June 10). Trends in the Development of Phishing and Countermeasures. https://my-itspecialist.com/trends-in-phishing-development-and-countermeasures.
11. CyberSecureFox. (2024, November 15). Sharp Increase in Phishing Attacks on the Financial Sector: New Cybersecurity Challenges in 2024. https://cybersecurefox.com/uk/zrostannya-fishingovyh-atak-finansovyi-sektor-2024/
12. SlashNext. (2024). The State of Phishing 2023. Pleasanton, CA: SlashNext Security Products. https://slashnext.com/wp-content/uploads/2023/10/SlashNext-The-State-of-Phishing-Report-2023.pdf
13. 10Guards. (2024, September 23). Surge in DDoS Attacks by 46% in the First Half of 2024. *Cybersecurity Services (Advisory/Ethical Hacking)*. https://10guards.com/ua/blog/2024/09/23/surge-in-ddos-attacks-gcore-report-reveals-46-increase-in-first-half-of-2024.
14. *GitHub*. (2015). Jehuty4949. NSL-KDD Dataset. https://github.com/Jehuty4949/NSL_KDD.