УДК 004.056.53

А.А. Булаївський

Н.Р. Кондратенко

ФІШИНГ: МЕТОДИ ТА ЗАСОБИ ЗАХИСТУ НА ОСНОВІ МОДЕЛЕЙ РЕГРЕСІЙНОГО АНАЛІЗУ

Вінницький національний технічний університет

Анотація. *У роботі досліджено прогнозування інтенсивності фішингових атак на основі кореляційно-регресійного аналізу та навчання за допомогою нейромереж. Визначено вплив соціально-економічних факторів на динаміку загроз і побудовано математичну модель, точність якої підтверджено статистичними метриками. Описано застосування нейромережевих архітектур для ідентифікації складних форм соціальної інженерії. Обґрунтовано доцільність використання гібридних підходів для прискорення реакції на інциденти порівняно з традиційними засобами захисту.*

Ключові слова: фішинг, регресія, соціальна інженерія, кібербезпека, нейронні мережі.

Abstract. *The paper examines phishing intensity prediction based on correlation-regression analysis and learning using neural networks. The impact of socio-economic factors on threat dynamics is determined, and a mathematical model is developed, with its accuracy verified by statistical metrics. The application of neural network architectures for identifying complex forms of social engineering is described. The study justifies the use of hybrid approaches to accelerate incident response compared to traditional security tools.*

Keywords: phishing, regression, social engineering, cybersecurity, neural networks.

Вступ

Стрімкий розвиток методів соціальної інженерії та поява нових векторів атак, таких як BEC та Quishing, роблять фішинг однією з головних загроз для сучасної IT-інфраструктури. Традиційні засоби захисту часто виявляються неспроможними протидіяти складним персоналізованим атакам, що зумовлює потребу у впровадженні інтелектуальних систем аналізу [1]. Застосування кореляційно-регресійних моделей у поєднанні з навчанням за допомогою нейромереж дозволяє здійснювати превентивний моніторинг загроз та забезпечує високу точність ідентифікації шкідливого контенту в режимі реального часу.

Результати дослідження

Аналіз сучасних методів фішингу дозволив виокремити основні типи атак у 2024-2025 роках. Email phishing представляє масові розсилки з підробленими повідомленнями від імені банків або компаній з метою отримання облікових даних. Spear phishing характеризується цільовими атаками з використанням персоналізованої інформації про жертву. Whaling спрямовано на керівників вищої ланки для отримання доступу до конфіденційної інформації. BEC (Business Email Compromise) реалізується через скомпрометовані легітимні облікові записи, що дозволяє обходити класичні методи автентифікації SPF, DKIM та DMARC. Quishing використовує QR-коди для приховування шкідливих URL у графічних об'єктах, що робить їх невидимими для традиційних фільтрів [2].

Система протидії фішингу базується на багаторівневому підході: технічний рівень включає протоколи автентифікації та фільтрацію на поштових шлюзах, організаційний рівень передбачає багатофакторну автентифікацію та обмеження прав доступу, людський фактор вимагає навчання персоналу розпізнаванню ознак фішингу, аналітичний рівень забезпечується системами поведінкової аналітики для виявлення аномалій у стилістиці повідомлень [3, 4].

Приклад BEC-атаки у березні 2024 року демонструє типовий сценарій: зловмисники отримали доступ до пошти фінансового директора через credential stuffing, протягом п'яти днів вивчали стиль

комунікації та поточні проєкти компанії, після чого надіслали бухгалтеру лист з терміновим проханням здійснити переказ коштів. Лист пройшов усі перевірки SPF/DKIM, а стилістика відповідала звичайному стилю CFO. Збитки склали 127,000 доларів, атаку виявлено через аномалію в поведінці керівника [4].

Для аналізу тенденцій розвитку фішингових атак використано лінійну модель регресійного аналізу. Дані про кількість фішингових атак отримано з квартальних звітів Anti-Phishing Working Group (APWG) за період з першого кварталу 2022 року по другий квартал 2023 року [2]. Для визначення факторів, що впливають на інтенсивність атак, проаналізовано три соціально-економічні показники з офіційних джерел: обсяг заборгованості за кредитними картками, темпи зростання ВВП США та обсяги електронної комерції. Дані представлено у таблиці 1.

Таблиця 1 – Вихідні дані для кореляційно-регресійного аналізу

Квартал	Кількість фішингових атак (APWG)	Заборгованість за кредитними картками, млрд %	Зростання ВВП, %	Обсяги електронної комерції, млрд $
Q1 2022	1025968	840	-1	287.9
Q2 2022	1097811	890	0.3	292.6
Q3 2022	1270883	930	2.7	300.9
Q4 2022	1350037	986	3.4	308.9
Q1 2023	1624144	986	2.8	300.7
Q2 2023	1286208	1030	2.4	306.7

Для оцінки сили зв'язку між кожним фактором та кількістю фішингових атак застосовано коефіцієнт кореляції Пірсона, який розраховується за формулою: $r = \frac{\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})}{S_x * S_y}$, де $\bar{x}$ та $\bar{y}$ – середні значення змінних, S_x та S_y – їх стандартні відхилення [6]. Кореляція вимірює ступінь залежності між двома змінними, а регресія дозволяє побудувати математичну модель, яка описує залежність між незалежною змінною та залежною змінною [3]. Результати кореляційного аналізу представлено у таблиці 2.

Таблиця 2 – Коефіцієнти кореляції між фішинговими атаками та соціально-економічними факторами

Фактор	Коефіцієнт кореляції r	Джерело
Заборгованість за кредитними картками	0.801	NY Fed Consumer Credit Panel
Темпи зростання ВВП США	0.794	US Bureau of Economic Analysis
Обсяги електронної комерції	0.609	US Census Bureau

Найвищий показник кореляції отримано між обсягом заборгованості за кредитними картками та кількістю фішингових атак (r = 0.801), що вказує на сильний прямий лінійний зв'язок. Це узгоджується з гіпотезою про те, що зростання використання кредитних карток для онлайн-платежів створює більше можливостей для фішингових атак на платіжну інформацію користувачів. Темпи зростання ВВП також продемонстрували сильну кореляцію (r = 0.794), що відображає загальну економічну активність та інтенсивність онлайн-транзакцій. Обсяги електронної комерції показали помірну кореляцію (r = 0.609), що може пояснюватися впливом сезонних факторів на онлайн-торгівлю.

На основі найсильнішої кореляції побудовано модель регресійного аналізу вигляду: $y = a + bx$, де x – обсяг заборгованості за кредитними картками (незалежна змінна), y – кількість фішингових атак (залежна змінна). Для розрахунку коефіцієнтів регресії використано метод найменших квадратів. Середні значення становлять: $\bar{x}$ = 943.67, $\bar{y}$ = 1275842.

Сума добутків відхилень: $\sum(x_i - \bar{x})(y_i - \bar{y}) = 54306409$.

Сума квадратів відхилень: $\sum(x_i - \bar{x})^2 = 24851.34$.

Коефіцієнт нахилу: $b = \frac{\sum x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{54306409}{24851.34} = 2185.1$.

Сталий параметр: $a = \bar{y} - b * \bar{x} = 1275842 - 2185.1 * 943.67 = -785803$.

Рівняння регресії має вигляд: $y = 2185.1x - 785803$, що означає зростання кількості фішингових атак на 2185 випадків при збільшенні заборгованості за кредитними картками на один мільярд доларів.. Коефіцієнт детермінації $R^2 = 0.642$ свідчить про те, що модель пояснює 64.2% відхилень даних [3].

Для оцінки точності моделі використано метрику RMSE (Root Mean Square Error):

$$RMSE = \sqrt{\frac{1}{n}\sum(\bar{y}_i - y_i)^2},$$

де $\bar{y}_i$ – прогнозоване значення; y_i – фактичне значення.
Прогнозовані значення за моделлю та фактичні дані наведено у таблиці 3.

Таблиця 3 – Порівняння прогнозованих та фактичних даних

Квартал	Фактичне значення	Прогнозоване значення	Відхилення
Q1 2022	1025968	1049681	-23713
Q2 2022	1097811	1159037	-61226
Q3 2022	1270883	1246340	24543
Q4 2022	1350037	1368507	-18470
Q1 2023	1624144	1368507	255637
Q2 2023	1286208	1464850	-178642

Розрахована похибка становить RMSE = 130715 атак, що складає 10.24% від середнього значення кількості атак. Отримане значення є прийнятним для прогнозування тенденцій у кібербезпеці та підтверджує придатність моделі для короткострокового планування ресурсів захисту [3].
Графічне представлення регресійної моделі демонструє залежність між обсягом заборгованості за кредитними картками та кількістю фішингових атак протягом досліджуваного періоду (рис. 1).

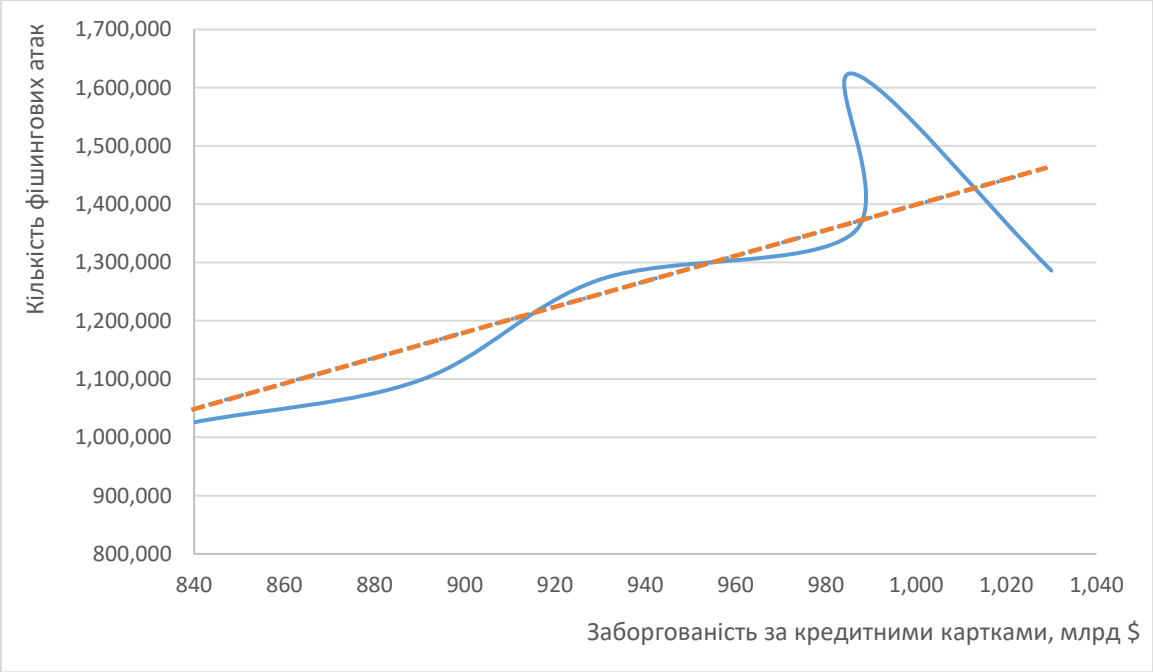

Рисунок 1 – Регресійна модель кількості фішингових атак та обсягом заборгованості за кредитними картками

Отримані результати підтверджують вплив обсягу використання кредитних карток на інтенсивність фішингових атак, що пояснюється зростанням кількості онлайн-транзакцій та привабливості платіжної інформації для зловмисників.

Для подолання загроз актуальні галузеві підходи до кібербезпеки передбачають застосування навчання за допомогою нейромереж, які демонструють значно вищу ефективність у виявленні фішингу [4, 5]. Згорткові нейронні мережі (CNN) використовують для аналізу візуальних елементів та виявлення підроблених логотипів, рекурентні мережі (RNN/LSTM) аналізують текстові закономірності для розпізнавання соціальної інженерії, а трансформери на основі BERT та GPT забезпечують контекстний аналіз з виявленням маніпулятивних фраз [5, 6]. Порівняльний аналіз показує, що сигнатурний аналіз досягає точності 76%, лінійна регресія – 87%, а методи глибокого навчання забезпечують показник правильності 96%, при цьому точність (precision) 94% та повнота (recall) 95% [6]. Оптимальний підхід

поєднує обидва методи в гібридній архітектурі, де модель парної регресії використовується для прогнозування масових хвиль атак на основі соціально-економічних факторів, а deep learning забезпечує точний аналіз окремих повідомлень, що дозволяє скоротити час від виявлення до блокування на 73% порівняно з традиційними способами захисту [6]. Побудована модель може бути використана для планування ресурсів систем кібербезпеки та розробки стратегій протидії фішингу з урахуванням динаміки фінансових показників.

Висновки

Проведене дослідження підтвердило, що аналіз соціально-економічних чинників є ключовим для прогнозування інтенсивності фішингу. Сильний прямий зв'язок між заборгованістю за кредитними картками та кількістю атак (r = 0.801) дозволив побудувати регресійну модель, яка пояснює 64,2% дисперсії даних. Це робить її ефективним інструментом для стратегічного планування ресурсів кіберзахисту та виявлення масових хвиль загроз.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Баришев Ю., Кондратенко Н., Казміревський В., Кирилащук Т. Нечіткі множини типу-2 в задачах моделювання та оцінювання станів критичних систем з недовизначеними вхідними даними та використанням експертів // Інформаційні технології та комп'ютерна інженерія. – 2023. – Том 57, № 2. – С. 13–24. Режим доступу: https://itce.vn.ua/uk/journals/t-57-2-2023 (дата звернення: 04.01.2026).
2. APWG Phishing Activity Trends Report [Електронний ресурс] // Anti-Phishing Working Group. – 2022-2023. – Режим доступу: https://apwg.org/trendsreports/ (дата звернення: 03.01.2026).
3. Statistical methods for predicting phishing attacks [Електронний ресурс] // ResearchGate. – 2024. – Режим доступу: https://www.researchgate.net/publication/379629280_STATISTICAL_METHODS_FOR_PREDICTING_PHISHING_ATTACKS (дата звернення: 03.01.2026).
4. What Is Business Email Compromise (BEC)? Tactics and Prevention [Електронний ресурс] // Palo Alto Networks. – 2024. – Режим доступу: https://www.paloaltonetworks.com/cyberpedia/what-is-business-email-compromise-bec-tactics-and-prevention (дата звернення: 03.01.2026).
5. A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN [Електронний ресурс] / K. Alkawaz, S. Steven, A. Hajamydeen // MDPI: Electronics. – 2023. – Т. 12, № 1. – С. 232. – Режим доступу: https://www.mdpi.com/2079-9292/12/1/232 (дата звернення: 03.01.2026).
6. Meléndez R. Comparative Investigation of Traditional Machine Learning Models and Transformer Models for Phishing Email Detection [Електронний ресурс] / R. Meléndez, M. Ptaszynski, F. Masui // Preprints.org. – 2024. – Режим доступу: https://www.preprints.org/manuscript/202410.1467/v1 (дата звернення: 03.01.2026).

Булаївський Артем Андрійович – студент 3 курсу групи 2БС-23Б, факультет інформаційної технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, email: artembulaivskiy@gmail.com.

Кондратенко Наталія Романівна – к.т.н., проф., професор кафедри захисту інформації, Вінницький національний технічний університет, м. Вінниця, email: kondratenko.natalia@vntu.edu.ua.

Bulaivskyi Artem Andriyovych – third-year student of group 2BS-23B, Faculty of Information Technology and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, email: artembulaivskiy@gmail.com.

Kondratenko Nataliia Romanivna – Phd in Technical Sciences, Professor, Department of Information Security, Vinnytsia National Technical University, Vinnytsia, email: kondratenko.natalia@vntu.edu.ua.