УДК 004.056

Боднар І.І.

АНАЛІЗ ФІШИНГОВИХ АТАК НА ОСНОВІ КОРЕЛЯЦІЙНО - РЕГРЕСІЙНОЇ МОДЕЛІ

Вінницький національний технічний університет

Анотація

У статті досліджено сучасну проблематику фішингових атак, їх різновиди, а також використання кореляційно-регресійних моделей для аналізу і прогнозування тенденцій розвитку цих атак. Описано методи оцінки залежностей між кількістю виявлених фішингових сайтів та соціально-економічними факторами. Визначено наближеність даних побудованої регресійної моделі оцінюється за допомогою метрики RMSE.

Ключові слова: фішинг, кореляція, регресія, прогнозування.

Abstract

The article examines the modern issues of phishing attacks, their types, and the use of correlation-regression models for analyzing and predicting trends in their development. Methods for assessing dependencies between the number of detected phishing websites and socio-economic factors are described. The approximation of the data of the constructed regression model is determined using the RMSE metric.

Keywords: phishing, correlation, regression, prediction.

Сучасна проблематика фішингових атак полягає в тому, що зловмисники постійно вдосконалюють свої методи. Вони використовують персоналізацію, штучний інтелект і машинне навчання для створення реалістичних повідомлень, які важко відрізнити від справжніх. Особливу небезпеку становить таргетований фішинг, коли атаки спрямовані на конкретних осіб або організації з використанням зібраної інформації про жертву.

Додатковою проблемою є складність ідентифікації фішингових сайтів через технічну схожість із легітимними веб-ресурсами, включаючи підроблені доменні імена, шифрування та дизайн. Зростає також використання багатоканальних атак: одночасне залучення електронної пошти, соціальних мереж і мобільних додатків для введення користувачів в оману.

Ситуацію ускладнює поширення фішингових атак через корпоративні канали, де підроблені повідомлення імітують внутрішні запити від IT-відділу або HR. Це посилює ризики для компаній, які можуть втратити як дані клієнтів, так і конфіденційну інформацію.

Суспільство недостатньо обізнане про нові види атак, що дозволяє зловмисникам використовувати страх, поспішність і необачність користувачів для досягнення своїх цілей. Потреба у безперервному навчанні та інформуванні залишається актуальною, оскільки методи атак адаптуються швидше, ніж захисні механізми.

Фішинг має різноманітні форми, залежно від методів і каналів, які використовують зловмисники. Класичний фішинг базується на розсилці електронних листів, що імітують повідомлення від банків, платіжних систем або популярних сервісів, з метою отримання конфіденційної інформації. Спрямований фішинг або спірфішинг є більш персоналізованим і орієнтується на конкретну жертву чи організацію, використовуючи дані, зібрані заздалегідь.

Фармінг спрямований на перенаправлення користувачів на підроблені вебсайти через маніпуляцію DNS-записами або встановлення шкідливого програмного забезпечення. SMS-фішинг, також відомий як смшинг, реалізується через текстові повідомлення, які спонукають користувачів переходити за підозрілими посиланнями. Вішинг використовує телефонні дзвінки, де шахраї видають себе за представників банків або інших організацій, щоб виманити конфіденційну інформацію.

Фішинг у соціальних мережах передбачає створення підроблених профілів або поширення фальшивих конкурсів, які спонукають до передачі особистих даних. Clone-фішинг реалізується через копіювання реальних повідомлень із додаванням шкідливих файлів або посилань. Кіт-фішинг орієнтується на високопосадовців або керівників, використовуючи їхню важливість у компанії як перевагу для атак.

Окрім цих форм, фішинг також активно використовує маніпуляції в корпоративному середовищі, де підроблені запити можуть виглядати як внутрішні повідомлення від IT-відділу. У сучасних реаліях

зловмисники також експериментують із використанням зловмисних рекламних оголошень для введення користувачів в оману.

Кореляційно-регресійна модель прогнозування є одним з основних методів аналізу та прогнозування трендів кібератак, що використовує взаємозв'язки між різними змінними для побудови прогнозу.

Кореляція вимірює ступінь залежності між двома змінними[2]. Висока кореляція означає, що зміни в одній змінній можна передбачити за змінами іншої змінної. Це дозволяє зрозуміти, як зміна однієї характеристики може вплинути на іншу.

Регресія йде далі і дозволяє побудувати математичну модель, яка описує залежність між незалежними змінними (факторами) і залежною змінною (результатом). Це дозволяє робити точніші прогнози, оскільки регресія не лише виявляє взаємозв'язки, а й дає змогу моделювати їх чисельно.

Лінія регресії будується шляхом знаходження математичного виразу, який описує залежність між змінними. Спочатку збирається набір даних із двома змінними: незалежною, яка визначає вплив, та залежною, що відображає результат. Основне завдання — знайти пряму, яка мінімізує відхилення фактичних значень залежної змінної від значень, передбачених цією прямою.

Для цього використовується метод найменших квадратів, який обчислює параметри рівняння прямої. Рівняння має вигляд

$$y = ax + b$$

де b — точка перетину з віссю y

a — коефіцієнт нахилу, що показує, як зміна незалежної змінної впливає на залежну.

Обчислення параметрів базується на аналізі середніх значень змінних, а також на кореляції між ними. Після визначення коефіцієнтів лінія проводиться через простір даних, показуючи найкращу відповідність між змінними.

Проведемо аналіз даних фішингових атак. Згідно звіту Google у 2020 році було виявлено більш ніж два мільйони фішингових сайтів, що становить приблизно 40 тисяч випадків на тиждень (рис.1).

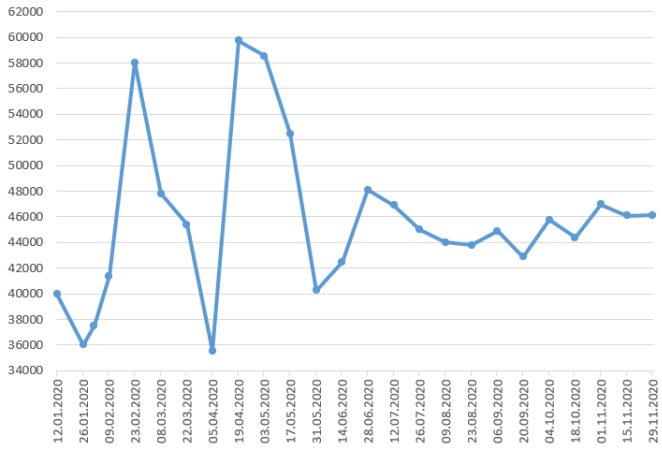

Рисунок 1 — Статистика виявлених фішингових сайтів

Для того, щоб визначити основні тенденції розвитку кібератак, необхідно перевірити кореляцію між кількістю виявлених атак і чинниками (факторами), що імовірно можуть позначатися на коливаннях інтенсивності атак.

запропоновано внести наступні фактори соціальні та соціально-економічні фактори:

C_1 – Відсоток безробіття в США;

C_2 – Індекс споживчих цін;

C_3 – Захворюваність на COVID-19;

C_3 –Середня з/п в США;

C_3 –Середня з/п в Україні;

C_3 –Заборгованість з/п в Україні;

Результати обрахунків зведено в таблицю 1

Таблиця 1 — Таблиця коефіцієнтів кореляції парної моделі між даними фішингової атаки та соціальними та соціально-економічними факторами

Фактор	Значення кореляції
Безробіття в США(%)	0,636250848
Індекс споживчих цін	-0,715016857
Середня з/п в США	0,569566564
Середня з/п в Україні	-0,23266
Заборгованість з/п в Україні	-0,155631015
Захворюванність на COVID-19	0,932239532

Найвищий показник коефіцієнта кореляції отримано між даними фішингових сайтів та захворюваність на COVID-19.

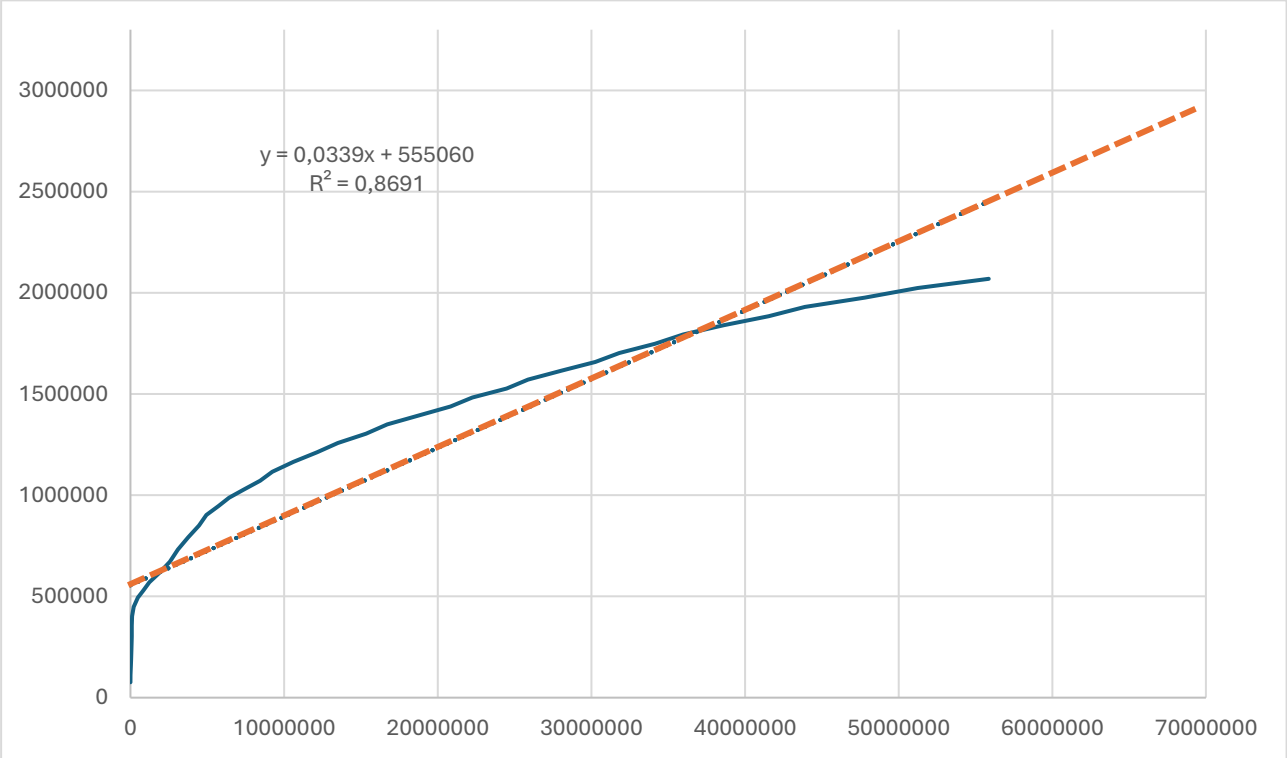

Рисунок 2 — Регресійна модель кількості виявлених фішингових сайтів та кількістю захворівших на COVID-19

Для оцінки наближення даних важливо використовувати кількісні метрики, які надають зрозуміле уявлення про те, наскільки добре модель відповідає фактичним даним. Однією з таких ключових метрик є корінь середньоквадратичної помилки (RMSE). Цей показник вимірює середнє відхилення прогнозованих значень моделі від реальних, що дозволяє оцінити загальний рівень помилки. Використання RMSE є доцільним, оскільки він чутливий до великих відхилень, акцентуючи увагу на помилках, які можуть суттєво вплинути на якість моделі. Ця метрика допомагає не лише оцінити ефективність моделі, але й порівняти її з іншими моделями для вибору оптимального підходу.

RMSE обраховується за формулою:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}},$$

де $\hat{y}_i$ — прогнозоване значення;

y_i — спостережене значення;

n — кількість елементів вибірки.

Проведені розрахунки на основі спостережених значень дали результат: RMSE=214742≈19.62%

Висновки

Розглянуто сучасні проблеми, пов'язані з фішинговими атаками, їх видами та методами аналізу на основі кореляційно-регресійних моделей. Основною метою дослідження було визначення залежностей між кількістю виявлених фішингових сайтів та соціально-економічними факторами, які можуть впливати на інтенсивність атак.

Отримані результати показали, що існує вплив на кількість фішингових атак захворюваність на COVID-19, що підтверджується найвищим з розглянутих значенням коефіцієнта кореляції (0,932). Це свідчить про сильний взаємозв’язок між цими змінними та демонструє, як глобальні соціальні кризи можуть впливати на кіберзагрози. Інші фактори, такі як рівень безробіття та середня заробітна плата, також показали певний вплив, але їх значення кореляції були помірними або низькими.

Для моделювання взаємозв’язків була використана кореляційно-регресійна модель, яка дозволяє виявити кількісні залежності між змінними. Оцінка похибки наближення за допомогою RMSE показала, що середня похибка прогнозу становить 19,62%, що є прийнятним рівнем для аналізу такого типу фішингових атак. Ця метрика показує що в умовах зібраних даних за фішингових атак є перспектива створювати прогнозні моделі в умовах невизначеності на основі нечіткої логіки.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Алексєєнко О. А. Технології фішингу в мобільних системах. *Сучасна інформаційна безпека* . 2022. Вип. 1, № 49. URL: https://doi.org/10.31673/2409-7292.2022.014449 (дата звернення: 10.12.2024).
2. Регресійний аналіз тенденцій розвитку кібератак С. О. Аксьончиков, І. В. Ємельянова, К. Д. Маркова, І. І. Сватовський
3. Міжнародна організація праці. *ILOSTAT database* [Електронний ресурс]. Режим доступу: https://ilostat.ilo.org/data/ (дата звернення: 10.12.2024).
4. ITWiki. *Root Mean Square Error* [Електронний ресурс]. – Режим доступу: https://itwiki.dev/data-science/ml-reference/ml-glossary/root-mean-square-error (дата звернення: 10.12.2024). – Назва з екрана.

Боднар Ілля Іванович — студент групи 1БС-23м, факультет інформаційних технологій та комп’ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail:illiabodnar2017@gmail.com

Bodnar Illia I.— student of group 1BS-23m, Faculty of Information Technology and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: :illiabodnar2017@gmail.com

Кондратенко Наталя Романівна— к. т. н., професор кафедри захисту інформації, Вінницький національний технічний університет, м. Вінниця

Kondratenko Natalia Romanivna — PhD (eng), professor of information protection department, Vinnytsia National Technical University, Vinnytsia